Renewable and Sustainable Energy Reviews 27 (2013) 362–400

Contents lists available at SciVerse ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

A review of measure-correlate-predict (MCP) methods used to estimate long-term wind characteristics at a target site

José A. Carta [a,*], Sergio Velázquez [b], Pedro Cabrera [a]

[a] Department of Mechanical Engineering, University of Las Palmas de Gran Canaria, Campus de Tafira s/n, 35017 Las Palmas de Gran Canaria, Canary Islands, Spain
[b] Department of Electronics and Automatics Engineering, University of Las Palmas de Gran Canaria, Campus de Tafira s/n, 35017 Las Palmas de Gran Canaria, Canary Islands, Spain

ARTICLE INFO

Article history:
Received 17 April 2013
Received in revised form
2 July 2013
Accepted 5 July 2013
Available online 30 July 2013

Keywords:
Measure-correlate-predict method
Regression analysis
Data mining
Spatial correlation
Wind speed
Wind direction

ABSTRACT

So-called Measure-Correlate-Predict (MCP) methods have been extensively proposed in renewable energy related literature to estimate the wind resources that represent the long-term conditions at a target site where a short-term wind data measurement campaign has been held.

The main differences between the various MCP methods lie fundamentally in the type of relationship established between the wind data (speed and direction) recorded at the target site and the wind data recorded simultaneously at one or various nearby weather stations which serve as reference stations and for which long-term data series are also available.

This paper reviews a wide range of MCP methods that have been proposed in the context of wind energy analysis, a number of which have been implemented in wind energy industry software applications. This review includes the initial methods first proposed in the 1940s which generally attempted only to estimate the long-term mean annual wind speed from a single reference station, and extends up to the most recent methods proposed in the present century based on automatic learning techniques which use several reference stations.

In addition to offering a description of the linear, non-linear and probabilistic transfer functions used by the different algorithms, the hypotheses on which these functions are based and the data format with which the various methods work (time series or frequency distributions), this review will also cover limitations in the use of MCP methods, the uncertainty associated with them and the different reference data sources that have been studied. In this sense, the extensive collection of MCP methods which is brought together and reviewed in this paper, ranging from the simplest and easiest-to-use models to the most complicated computational ones which require specific user experience, comprises an extremely useful catalogue when it comes to choosing the best predictor method.

Contents

* Corresponding author. Tel.: +34 928 45 14 83; fax: +34 928 45 14 84.
E-mail address: jcarta@dim.ulpgc.es (J.A. Carta).

1364-0321/$ - see front matter
http://dx.doi.org/10.1016/j.rser.2013.07.004

1. Introduction

Due to the interannual variability of wind speed [1–3] various authors [2,5,6,9–16] have indicated that wind data measured over just a few years are insufficient to reflect the average conditions present over the lifetime of a wind project (typically 20 years).

To estimate the long-term wind resource at a candidate site a long series of wind data needs to be available for that site. There is no general agreement as to the recommended time period that needs to be covered by such wind data time series for a candidate site. Some authors suggest data availability is required for a three year period [12], some for a 10 year period [6,13], whilst others state that 20 or even 30 years are required to correctly characterise the long-term wind resource of a site [2,10,14].

As a result of the expense associated with carrying out measurement campaigns when no historical series of wind data are available at the target site and as a consequence of the expense caused by the delay [4,15,16] to the development of the wind farm, such measurement campaigns tend to be restricted to 12 months or less [17]. It is highly unlikely that such a short measuring period will provide results that reflect the long-term conditions [18].

A possible alternative approach [13] involves linking the candidate site with one or more nearby sites where long series of wind data measurements have been recorded. Such a link can be achieved by using both physical[1] and statistical methods. The latter have been extensively proposed in the scientific literature for this purpose [2,4–5,7,9,11–17,19–22] and are commonly known as Measure-Correlate-Predict (MCP) methods.

1.1. Basic idea behind the approach of MCP methods

The aim when using MCP methods is to perform long-term hindcasting of the wind conditions at a candidate site for which only short-term wind data series are available. This hindcasting is based on the use of historical and homogenous wind data series which are available for nearby weather stations used as reference stations. These methods also require part of the long-term data series from the reference stations to coincide in length and time period with the short-term data recorded at what is variously described in the scientific literature as the target site, candidate site or survey site. Given the higher frequency of use of the term 'target site', this description will be used hereafter in this review.

The block diagram seen in Fig. 1 shows the procedure normally employed by MCP methods. As can be seen, the procedure basically comprises two stages. In the first stage (indicated by an encircled number "1" in Fig. 1), using as a starting point the wind data series recorded at the reference and target sites for the short-term period that is common to both (concurrent data period), the aim is to establish a relationship between the two sets of data. In the second stage (indicated by an encircled number "2" in Fig. 1), the long-term wind data series available for the reference stations are applied as input variables to the relationship established in stage 1 in order to perform long-term hindcasting of the wind conditions at the target site. The few MCP methods which do not follow the procedure outlined in Fig. 1 determine the relationship between the short- and long-term data of the reference station and then apply it to the short-term data available for the target site with the aim of obtaining data that represent its long-term wind conditions.

1.2. Importance of MCP methods in wind resource estimation

According to Addison et al. [22], MCP methods tend to provide higher accuracy than physical modelling methods (for example, WAsP[2] [20,23]), especially in complex environments. Physical models also introduce non-quantifiable uncertainties in the prediction. For these reasons, MCP methods have evolved to become a standard tool for wind farm developers and have been implemented in wind energy industry software applications [24–30] for long-term wind speed forecasts. As reported in a study by Derrick et al. [31] in 1996, almost all the consultants who have carried out measurements to assess the wind regime at a site have used some form of measure-correlate-predict strategy to forecast the long-term wind regime. Sanz [32] presents the results from a wind resource assessment questionnaire called WAUDIT[3] which had been sent during the first half of 2010 to wind analysts ranging from industrial to academic sectors at European organisations operating in the wind energy field. The purpose of the questionnaire was to construct a knowledge map for wind resource assessment methods in the industry. A total of 72 wind analysts from 48 different organisations and 13 European countries contributed to the survey. To the surprise of the author of the report, despite the importance of MCP methods for wind energy developers, only 70% of all respondents (80% in the case of consultants) said that they used these methods.

[1] Numerical wind flow models based on theoretical approaches and which can be classified into categories of different degrees of complexity.

[2] WAsP (Wind Atlas Analysis and Application Program) is a model developed by the Wind Energy Department at Risø National Laboratory for wind resource assessment. WAsP is probably the most widely used numerical wind flow model in the wind energy industry.

[3] WAUDIT is an Initial Training Network (ITN) Marie–Curie action financed by the FP7-PEOPLE programme (Project Reference:238576) of the European Commission. Under the subject of "Wind Resource Assessment Audit and Standardization", the aim of WAUDIT is to set up a group of researchers in the field of wind resource assessment, for the purpose of supporting the technological development of wind energy as one of the world's fastest growing industries.

Nomenclature

A orthogonal matrix of L rows and L columns, formed by eigenvectors of the matrix product of $\boldsymbol{V}(L \times M)$ and its transpose $\boldsymbol{V}^T(M \times L)$;
A parameter defined by Eq. (76);
$a_{i,k}$ element of matrix **A**, Eq. (94);
a_0,a_k parameters of Eq. (5);
ANNs artificial neural networks;
AR matrix where each element $ar_{i,j}$ represents the average ratio corresponding to wind speed bin i of the wind direction sector j of the reference site;
$ar_{i,j}$ element of the matrix **AR**, Eqs. (68)–(70);
ASOS automated surface observing system;
B matrix in which an element $b_{i,j}$ represents the wind speed bin i and wind direction sector j of the reference site;
b wind speed bin, Eq. (29);
b_i Bin i of the wind speed, Eq. (18);
$b_{i,j}$ element of the matrix **B** which contains the set of ratios calculated between each target site wind speed, $(v_k)_t^{ST}$, which is paired by date and time k with the corresponding reference site wind speed, $(v_k)_r^{ST}$;
b_k parameter of Eq. (5);
BMARS Bayesian multivariate adaptive regression spline;
BNs Bayesian networks;
C matrix of coefficients $c_{i,j}$;
C Weibull scale parameter;
CDF cumulative distribution function;
CFSR climate forecast system reanalysis;
$c_{i,j}$ coefficients, Eq. (80), of matrix **C**, which assuming that they remain constant over the long-term, are used to estimate the long-term relative frequencies of occurrence of category i at the reference site and of category j at the target site, Eqs. (80) and (81);
CNF normalization factor, Eqs. (86)–(89);
CPMF cumulative probability mass function;
$\mathrm{CPMF_j}$ cumulative probability mass function of the target site wind speed for bin j of the reference site;
CSQ Chi-square regression;
$c(k)$ the simple autocovariance function, Eqs. (2) and (3);
C_r^{ST}, C_t^{ST} short-term Weibull scale parameters of the reference and target site, respectively, Eqs. (65) and (67);
d variable representing wind direction, Eq. (35);
$dd_{k,z}$ represents the distance between station k and station z, Eq. (101);
$(d_j)_r^{LT}$,$(d_k)_r^{LT}$ long-term reference site wind directions;
$(d_j)_r^{ST}$,$(d_k)_r^{ST}$ short-term reference site wind directions, Eq. (73);
$(d_k)_t^{ST}$ short-term target site wind directions, Eq. (73);
E square matrix of $N \times N$ wind direction sectors, Table 1;
E′ square matrix of $N \times N$ wind direction sectors whose elements are null or coincide with those of matrix **E** if the restriction imposed by Eq. (52) is met, Table 2;
ECMWF European centre for medium-range weather forecasts;
$e_{i,j}$ Element of matrix **E** which contains the number of times the wind has blown simultaneously in target site sector i and reference site sector j, Eq. (52);
$e'_{i,j}$ element of the matrix **E′**, Eqs. (52) and (53);
EWEA European wind energy association;
FFT fast Fourier transform;
$(F_i)_r^{LT}$, $(F_j)_t^{LT}$ long-term frequencies of occurrence of category i at the reference site and of category j at the target site, Eq. (81);
$(F_i)_r^{ST}$, $(F_j)_t^{ST}$ short-term frequencies of occurrence of category i at the reference site and of category j at the target site, Eq. (79);
$(f_{i,j})_r^{LT}$ long-term frequency of wind speed bin i and wind direction sector j at the reference site, Eqs. (17) and (18);
$(f_{i,j})_r^{ST}$ short-term frequency of wind speed bin i and wind direction sector j at the reference site, Eqs. (17) and (18);
$(f_j)_r^{LT}$ long-term wind speed frequency in wind direction sector j of the reference site, Eqs. (17)–(19);
$(f_j)_r^{ST}$ short-term wind speed frequency in wind direction sector j of the reference site, Eqs. (17)–(19);
$(FO_j)_t$ frequency of occurrence of WPT j at the target site t, Eq. (13);
$f_r^{LT}(v)$ long-term wind speed probability density function of the reference site, Eq. (86);
$f_r^{LT}(\theta_r)$ long-term wind speed frequency at the reference site for direction sector θ_r, Eq. (92);
$f_r^{ST}(\theta_r)$ short-term wind speed frequency at the reference site for direction sector θ_r, Eq. (92);
$f_t^{LT}(v)$ long-term wind speed probability density function of the target site, Eq. (86);
$f_t^{LT}(\theta_r)$ long-term wind speed frequency at the target site for direction sector θ_r, Eq. (92);
$f_t^{ST}(v_t)$ short-term wind speed probability density functions of the target site, Eq. (83);
$f_t^{ST}(v_t|v_r)$ wind speed probability density function at the target site conditioned on the reference site wind speed, Eq. (83);
$f_t^{ST}(\theta_r)$ short-term wind speed frequency at the target site for direction sector θ_r, Eq. (92);
$F_t^{ST}(v_t)$,$F_r^{ST}(v_r)$ cumulative distribution functions of wind speed at the target and reference site, respectively, Eqs. (83)–(85);
$F(v_t, v_r)$ bivariate Weibull distribution function, Eq. (90);
$f[\cdot]$ symbol representing a first-order linear function, Eq. (5);
G matrix where each element or cell, $g_{i,j}$, corresponds to a wind speed bin i (with default width of 1 m/s) and wind direction sector j (default value of 30°) of the reference site;
GCM general circulation model;
$g_{i,j}$ element of the matrix **G**;
GOR general orthogonal regression method;
H square matrix ($Nh \times Nh$) which constitutes a joint relative frequency distribution of short-term wind speed, wind direction and atmospheric stability;
$h_{i,j}^{ST}$ element of matrix **H** which represents the short-term relative frequency of occurrence of category i at the reference site and category j at the target site, Eqs. (79) and (80);
ITC installations of the Canary Technological Institute (Spanish initials);
ITN initial training network;
K Weibull shape parameter;
Kr number of time lags considered, Eqs. (2) and (3);
K_r^{ST}, K_t^{ST} short-term Weibull shape parameters of the reference site and target site, respectively, Eqs. (65) and (67);
L number of data of a wind speed time series, Eq. (94);
LT long-term;

M	number of parameters,a_ky b_k, in Eq. (5). Span of data, Eqs. (31) and (32). Number of meteorological stations, Eq. (94);
MCP	measure-correlate-predict;
MCRT	multiple climatic reduction technique;
MERRA	modern era retrospective-analysis for research and applications;
MMCP	multivariate MCP methodology;
MSD	moulded site data method or windscale method;
MTS	matrix time series method;
N	number of wind direction sectors under consideration, Eq. (52);
n	number of wind speed data under consideration, Eqs. (1), (3), (4), (23), (27), (31), (32), and (40);
Na	number of eigenvectors that need to be chosen such that Eq. (95) conserves a significant percentage of the variance expressed by the trace of $\mathbf{VV}^T$. $Na <$ minimum $\{L, M\}$, Eqs. (95)–(97);
NB	number of wind speed bins, Eqs. (17) and (18);
NCAR	National Centre for Atmospheric Research;
NCEP	National Centres for Environmental Prediction;
Nh	number of columns (or rows) of matrix **H**, Eq. (79);
$n_{i,j}$	number of ratios stored in the element $b_{i,j}$ of matrix **B**, Eqs. (68) and (69);
NP	order of the polynomial, Eq. (74);
Nw, Nd	numbers of wind speed and wind direction bins of the joint probability mass function, Eq. (82);
NWP	numerical weather prediction;
OLR	ordinary linear regression;
OR	orthogonal regression;
Pdf	probability density function;
PMF	probability mass function;
PMF_j	probability mass function of the target site wind speed conditioned on the reference site wind speed of bin j;
$p_r^{LT}(v,d)$	probability mass function of the long-term reference site wind speed and direction, Eq. (82);
$p_t^{LT}(v,d)$	probability mass function of the long-term target site wind speed and direction, Eq. (82);
$p_{t-r}^{ST}(v_t,d_t,v_r,d_r)$	joint probability mass function of the short-term wind speeds (v_t,v_r) and directions (d_t,d_r) of the target (t) and reference (r) sites, Eq. (82);
$p_t^{ST}(v_i,d_j\|v_k,d_z)$	probability density function of the target site wind speed conditioned on the reference site wind speed, Eqs. (88) and (89);
$p^{LT}[(\theta_j)_t]$	long-term probability that the wind blows in the direction corresponding to sector j of the target site (t), Eqs. (9), (10), and (12);
$p^{LT}[(\theta_k)_r]$	long-term probability that the wind blows in the direction corresponding to sector k of the reference site (r), Eqs. (9), (10) and (12);
$p^{ST}[(\theta_j)_t\|(\theta_k)_r]$	short-term probability that the wind blows in the direction corresponding to sector j of the target site (t), conditioned on the short-term probability that the wind blows in the direction corresponding to sector k of the reference site (r), Eqs. (9), (10), and (12);
$p^{ST}[\cdot\|\cdot]$	observed conditional probability at the target site during the short-term measurement period;
QR	quantile regression;
$\mathbf{Q}_t^T$	transpose of the column vector, $z=t$, of the square matrix **R**, Eq. (98);
R	square matrix of M rows and M columns, whose elements are defined by Eq. (101);
r	Pearson product-moment coefficient, Eq. (1). $-1\le r\le 1$;
Rf_j	Ratio between the long- and short-term wind frequency of wind direction sector j of the reference site, Eqs. (19) and (20);
RES	Renewable Energy Systems Ltd.
$r_{k,z}$	correlation coefficient between the data of station k and of station z, Eq. (101);
$r(k)$	cross-correlation coefficient as a function of time lag k, Eq. (2);
r_s	Spearman's correlation coefficient;
r^{LT}	correlation coefficient between the long-term wind speeds measured at the reference and target sites, Eq. (7);
Rv_j	Ratio between the long- and short-term mean wind speed in wind direction sector j of the reference site, Eqs. (19) and (20);
$[r^{ST}]_{(\theta_k)_r}$	correlation coefficient between the short-term wind speeds measured at the reference and target sites when the reference site wind blows in the direction corresponding to sector k, Eq. (11);
$r_{\Delta v \Delta d}$	correlation coefficient between the variables Δv and Δd, defined in Eqs. (72) and (73);
r1,r2	reference stations 1 and 2, Eq. (93);
R^2	coefficient of determination, Eq. (4);
SD	matrix in which each element, $sd_{i,j}$, represents the standard deviation of the ratios of wind speed bin i and wind direction sector j of the reference site;
$(s_i)_r^{ST}$	standard deviation of the ith short-term reference site wind speed, Eq. (27);
$(s_i)_t^{ST}$	standard deviation of the ith short-term target site wind speed, Eq. (27);
$sd_{i,j}$	element of matrix **SD**, Eqs. (69) and (71);
SOR	simple orthogonal regression method;
$SP_{i,j}$	estimated statistical property (mean standard deviation, correlation coefficient) in cell $g_{i,j}$ of matrix G, Eq. (74);
s_r^{LT}	standard deviation of observed long-term wind speeds at the reference site, Eq. (7);
s_r^{ST}	standard deviation of observed short-term wind speeds at the reference site, Eqs. (2) and (34);
s_r^{LT-ST}	standard deviation of the long-term (LT) mean wind speeds at the reference site averaged over periods as long as the short-term (ST) target period, Eq. (8);
$[s_r^{ST}]_{(\theta_k)_r}$	standard deviation of the short-term wind speeds recorded at the reference site when the wind at that site blows in the direction corresponding to sector k, Eq. (11);
ST	short-term;
s_t^{LT}	standard deviation of observed long-term wind speeds at the target site, Eqs. (7) and (8);
s_t^{ST}	standard deviation of observed short-term wind speeds at the target site, Eqs. (2), (8) and (34);
$[s_t^{ST}]_{(\theta_k)_r}$	standard deviation of the short-term wind speeds recorded at the target site when the wind at the reference site blows in the direction corresponding to sector k, Eq. (11);
$s_{v_r v_t}$	covariance between the $(v_j)_r^{ST}$ and the $(v_j)_t^{ST}$, Eqs. (23) and (25). Covariance of the target and reference site wind speeds, Eq. (44);
s_{v_r}	standard deviation of the reference site wind speed, Eqs. (42), (46), (48) and (49);
s_{v_t}	standard deviation of the target site wind speed, Eqs. (43), (46), (48) and (49);
$s_{v_r}^2$	variance of the wind speeds $(v_j)_r^{ST}$, Eqs. (23), (25) and (42);

$s^2_{v_t}$ variance of the $(v_j)_t^{ST}$, Eqs. (25) and (43);
$s_{(v_x)r}$ standard deviation of the X-axis components of the reference site wind speeds, Eq. (42);
$s_{(v_x)_r(v_x)_t}$ covariance of the X-axis components of the target and reference site wind speed, Eqs. (44) and (48);
$s_{(v_x)t}$ standard deviation of the X-axis components of the target site wind speeds, Eq. (43);
$s_{(v_x)_r(v_y)_t}$ covariance of the X-axis components of the reference site wind speed and the Y-axis components of the target site wind speed, Eqs. (44) and (49);
$s_{(v_y)r}$ standard deviation of the Y-axis components of the reference site wind speeds, Eq. (42);
$s_{(v_y)_r(v_x)_t}$ covariance of the Y-axis components of the reference site wind speed and the X-axis components of the target site wind speed, Eqs. (44) and (49);
$s_{(v_y)t}$ standard deviation of the Y-axis components of the target site wind speeds, Eq. (43);
$s_{(v_y)_r(v_y)_t}$ covariance of the Y-axis components of the target and reference site wind speed, Eqs. (44) and (48);
S1,S2 parameters defined by Eqs. (84) and (85);
$s_{\Delta v}$,$s_{\Delta d}$ standard deviations of the variables Δv and Δd defined in Eqs. (72) and (73);
$s^2_{\varepsilon v_t}$ variance of the errors of the dependent variable (target site wind speed), Eq. (24);
$s^2_{\varepsilon v_r}$ variance of the errors of the independent variable (reference site wind speed), Eq. (24);
$(t_k)_t^{LT}$ scaled time interval of long-term target site wind data, Eq. (20);
$(t_k)_t^{ST}$ time interval of short-term target site wind data measurement, Eq. (20);
$T1$,$T2$ ends of the interval of available wind data of a time series. $T1\geq 1$,$T2\leq L$,$T2-T1+1<N$, Eqs. (96), (97) and (99);
U matrix in which an element $u_{i,j}$, contains the number of times that, during the concurrent data period, target site wind speeds of bin i were recorded and reference site wind speeds of bin j were recorded;
U Random value extracted from a uniform distribution in the interval (0,1), Eq. (71);
$u_{i,j}$ element of the matrix **U**;
V matrix whose columns are time series with L data from M weather stations;
v variable representing wind speed, Eq. (35);
$v_{cut-off}$ reference-site cut-off speed, Eq. (30);
$v_{i,j}$ element of matrix **V**, Eq. (94);
$(v_i)_r^{ST}$,$(v_j)_r^{ST}$ short-term wind speeds observed at the reference site, Eqs. (23), (27), (31) and (33);
$(v_i)_t^{LT}(v_j)_t^{LT}$,$(v_k)_t^{LT}$ long-term wind speeds estimated at the target site, Eqs. (14), (16), (20), (21), (29), (30), (33), (34), (45), (55);
$(v_i)_t^{ST}$,$(v_k)_t^{ST}$,$(v_j)_t^{ST}$ short-term wind speeds observed at the target site, Eqs. (16), (22), (23), (27), (31) and (33);
$v_{i,t}$ target station wind speed (t), Eqs. (96) and (97);
$(v_j)_r^{LT}$ long-term wind speeds observed at the reference site, Eqs. (14), (21), (29), (30), (33) and (34);
$[(v_j)_r^{LT}]_{(\theta_k)_r}$ measured long-term wind speed at the reference site in function of wind direction sector k at the reference site, Eq. (15);
$\left[(v_j)_t^{LT}\right]_{(\theta_k)_r}$ estimated long-term wind speed at the target site in function of wind direction sector k at the reference site, Eq. (15);
$[(v_j)_x]_r^{LT}$ component X of the long-term reference site wind speed, Eqs. (36), (39) and (45);
$[(v_j)_x]_t^{LT}$ component X of the long-term target site wind speed, Eqs. (36), (38), (39),(45) and (102);
$[(v_j)_y]_r^{LT}$ component Y of the long-term reference site wind speed, Eqs. (37), (39) and (45);
$[(v_j)_y]_t^{LT}$ component Y of the long-term target site wind speed, Eqs. (37)–(39), (45), and (102);
v_{max} maximum wind speed, Eq. (62);
v_r variable representing the reference site wind speeds, Eq. (41);
$\mathbf{v}^{ST}$, $\mathbf{v}^{LT}$ vectors whose elements, $(\overline{v}_1^{ST},\ldots,\overline{v}_i^{ST},\ldots\overline{v}_M^{ST})^T$ and $(\overline{v}_1^{LT},\ldots,\overline{v}_i^{ST},\ldots\overline{v}_M^{LT})^T$ are the means of the square roots of the short-term, Eq. (99), and long-term, Eq. (100), wind speeds, respectively, calculated with the observed data of each of the M stations under consideration;
v_t variable representing the target site wind speeds, Eq. (41);
$(v_{tt})_r^{ST}$ reference site wind speed which is paired by date and time tt with the corresponding target site wind speed, $(v_{tt})_t^{ST}$;
$(v_{tt})_t^{ST}$ target site wind speed which is paired by date and time tt with the corresponding reference site wind speed, $(v_{tt})_r^{ST}$;
v_x variable representing the X-axis components of the wind speed, Eq. (35);
$(v_x)_r$ variable representing the X-axis components of the reference site wind speed, Eq. (41);
$(v_x)_t$ variable representing the X-axis components of the target site wind speed, Eq. (41);
v_y variable representing the Y-axis components of the wind speed, Eq. (35);
$(v_y)_r$ variable representing the Y-axis components of the reference site wind speed, Eq. (41);
$(v_y)_t$ variable representing the Y-axis components of the target site wind speed, Eq. (41);
$(\overline{v}_b)_r^{ST}$ mean short-term wind speed of bin b at the reference site, Eq. (29);
$(\overline{v}_b)_t^{ST}$ mean short-term wind speed of bin b at the target site, Eq. (29);
$(\overline{v}_i)_t^{LT}$ estimated long-term mean wind speeds in the target site wind direction sector i, Eqs. (55) and (56);
$(\overline{v}_j)_r^{LT}$ long-term mean wind speeds in the reference site wind direction sector j, Eqs. (18) and (19);
$(\overline{v}_j)_r^{LT}$ measured long-term mean wind speeds in the reference site wind direction sector j, Eqs. (55) and (56);
$(\overline{v}_j)_r^{ST}$ short-term mean wind speeds in the reference site wind direction sector j, Eqs. (18) and (19);
$(\overline{v}_x)_r^{ST}$ short-term mean wind speed of the X-axis wind component at the reference site, Eqs. (50) and (51);
$(\overline{v}_x)_t^{ST}$ short-term mean wind speed of the X-axis wind component at the target site, Eq. (50);
$(\overline{v}_y)_r^{ST}$ short-term mean wind speed of the Y-axis wind component at the reference site, Eqs. (50) and (51);
$(\overline{v}_y)_t^{ST}$ short-term mean wind speed of the Y-axis wind component at the target site, Eq. (51);
$\overline{v}_r^{LT}$ long-term mean wind speed measured at the reference site, Eqs. (6), (7) and (16);
$(\overline{v}_r^{LT})_j$ mean observed long-term wind speeds for WPT j at the reference site, Eq. (13);
$(\overline{v}_r^{ST})_j$ mean observed short-term wind speeds for WPT j at the reference site, Eq. (13);
$[\overline{v}_r^{LT}]_{(\theta_k)_r}$ measured long-term mean wind speed in wind direction sector k of the reference site, Eqs. (9) and (11);

$\bar{v}_r^{ST}$ short-term mean wind speed measured at the reference site, Eqs. (1), (3), (6), (7), (14), (16) and (34);
$[\bar{v}_r^{ST}]_{(\theta_k)_r}$ measured short-term mean wind speed of the reference site when the reference site wind speed blows in wind direction sector k, Eqs. (9), (11) and (15);
$\bar{v}_t^{LT}$ long-term mean wind speed estimated the target site, Eqs. (6), (7) and (13);
$[\bar{v}_t^{LT}]_{(\theta_j)_t}$ estimated long-term mean wind speed in wind direction sector j of the target site (t), Eqs. (9), (10) and (12);
$[\bar{v}_t^{LT}]_{(\theta_k)_r}$ estimated long-term wind speed at the target site for each given wind direction sector k at the reference site, Eqs. (11) and (12);
$\bar{v}_t^{ST}$ short-term mean wind speed measured at the target site, Eqs. (1), (3), (4), (6), (7),(14) and (34);
$(\bar{v}_t^{ST})_j$ mean observed short-term wind speeds for WPT j at the target site, Eq. (13);
$[\bar{v}_t^{ST}]_{(\theta_k)_r}$ measured short-term mean wind speed of the target site when the reference site wind speed blows in wind direction sector k, Eqs. (9), (11) and (15);
$(v'_k)_r^{ST}$ short-term reference site wind speeds smoothed out by use of the concept of moving averages, Eqs. (31) and (32);
$(v'_k)_t^{ST}$ short-term target site wind speeds smoothed out by use of the concept of moving averages, Eqs. (31) and (32);
$(\hat{v}_i)_t^{ST}$,$(\hat{v}_j)_t^{ST}$ estimated short-term wind speeds of the target site, Eqs. (4) and (5);
W Matrix of L rows and M columns, whose elements correspond to data projections of each station onto each row of **A** ($L \times L$);
WAsP Wind Atlas Analysis and Application Program;
WAUDIT an ITN Marie–Curie action financed by the FP7-PEOPLE programme (Project Reference: 238576) of the European Commission;
Weka Waikato Environment for Knowledge Analysis;
WF windiness Factor method;
$w_{k,j}$ element of matrix **W**, Eq. (94);
$w_{k,t}$ element of matrix **W**, corresponding to column $j=t$. 't' represents the target station, Eqs. (96) and (97);
WLS weighted total least squares;
WPT weather pattern type;
Z square matrix of $N \times N$, whose elements are expressed as a percentage of the total number data of each target station sector i, such that for each row (sector) i Eq. (54) is met, Table 1;
$z_{i,j}$ element of matrix **Z**, Eqs. (53), (55) and (56).

Greek letters

α offset of the regression line, Eqs. (21)–(23), (27), (30), (57),–(59), (61), (66), (67), (74) and (93);
α_x offset of the regression line of the X-axis components of the wind speeds, Eqs. (36), (45) and (50);
α_y offset of the regression line of the Y-axis components of the wind speeds, Eqs. (37), (45) and (51);
α_τ offset of the straight line for the quantile τ, Eq. (28);
α_1, α_2 offsets of the vector method, Eqs. (39) and (40);
β slope of the regression line, Eqs. (21), (22), (23), (27), (30), (33), (66) and (67);
β_{ij} slopes of the vector method, Eqs. (39) and (40);
β_j, β_i, α_j, α_i slope and offset, respectively, at the origin of the straight lines, Eqs. (55) and (56);
β_{kk} polynomial fit parameters, Eq. (74);
β_x slope of the regression line of the X-axis components of the wind speeds, Eqs. (36), (45), (46), (50) and (51);
β_y slope of the regression line of the Y-axis components of the wind speeds, Eqs. (37), (45), (47), (50) and (51);
β_1 slope of the regression line, Eqs. (57)–(61) and (93);
β_2 slope of the regression line, Eqs. (30), (57) and (93);
β_3 slope of the regression line, Eq. (32);
β_4 coefficient of interaction, Eqs. (57) and (59);
β_5 quadratic coefficient, Eqs. (58), (60) and (61);
β_6 cubic coefficient, Eq. (61);
β_{11}, β_{12}, β_{21}, β_{22} slopes of the vector method, Eq. (40);
β_τ slope of the straight line for the quantile τ, Eq. (28);
γ Weibull location parameter; parameter of the equation of fit between the data $r_{k,z}$ and $dd_{k,z}$, Eq. (101);
γ_r^{ST},γ_t^{ST} short-term Weibull location parameters of the reference and target site, respectively, Eq. (65);
ΔT_r temperature gradient or difference between temperatures of two heights at the reference site, Eq. (59);
Δv variable representing the differences between wind speeds (speed-up) at the target and reference sites, Eq. (75);
Δv_k differences between wind speeds (speed-up) at the target and reference sites, Eq. (72);
Δv^{LT} variable representing long-term differences between wind speeds (speed-up) at the target and reference sites, Eq. (77);
Δd variable representing the differences between wind direction (wind veer) at the target and reference sites;
Δd_k differences between wind direction (wind veer) at the target and reference sites, Eq. (73);
Δd^{LT} variable representing long-term differences between wind direction (wind veer) at the target and reference sites, Eq. (78);
δ Small fraction of the total number of data recorded in each wind direction sector which indicates the elements or bins which must be rejected, Eq. (52). Parameter of Eq. (63). Weibull shape factors ratio, Eqs. (66) and (67);
ε_j, ε_i white noise, Eqs. (21), (30), (58)–(61), (74) and (93);
$\varepsilon_{i,j}$ random variable with symmetric triangular distribution, Eqs. (70) and (71);
$\varepsilon_{i,t}^2$ mean square errors, Eq. (96);
$\varepsilon_{x,i}$ error associated with the X-axis component of the target site wind speed, Eq. (36);
$\varepsilon_{y,i}$ error associated with the Y-axis component of the target site wind speed, Eq. (37);
εv_r error associated with the reference site wind speed (independent variable), Eqs. (24) and (25);
εv_t error associated with the target site wind speed (dependent variable), Eqs. (24) and (25);
η parameter of the equation of fit between the data $r_{k,z}$ and $dd_{k,z}$, Eq. (101); parameter of Eqs. (62)–(64);
$(\theta_i)_r^{ST}$ short-term reference site wind direction sectors, Eq. (5);
$(\theta_j)_t$ wind direction sector j of the target station (t), Eq. (9);
$(\theta_k)_r$ wind direction sector k of the reference station (r), Eq. (9);
θ_r reference site wind direction sector;
λ Parameter which controls the degree of association between the variables v_s and v_r, Eq. (90). $0<\lambda\leq 1$. Represents scale parameter C or shape parameter K of the Weibull distribution, Eq. (91). Ratio between the variances of the errors of the dependent variable

(target site wind speed) and independent variable (reference site wind speed), Eqs. (24) and (25). Ratio between the sum of the variances of the errors of the dependent variable and of the equation and the variance of the independent variable, Eq. (26);

$\mu_{\Delta v}$,$\mu_{\Delta d}$ mean of the variables Δv and Δd, Eq. (75);

$(\rho_{v_r v_t})_x$ real component of the correlation vector, Eqs. (46) and (48);

$(\rho_{v_r v_t})_y$ imaginary component of the correlation vector, Eqs. (47) and (49);

$\rho_{v_r v_t}$ correlation vector between target and reference site wind speeds;

$\sigma_{\varepsilon Eq}$ error of the equation, Eq. (26);

τ quantile ($0 \leq \tau \leq 1$), Eq. (28);

χ^2 chi-square statistic, Eq. (27);

ψ coefficient of association, Eqs. (83)–(85).

Landberg and Mortensen [33] published a study in 1993 which, they declared, aimed to throw some light on the on-going argument between supporters of physical methods for wind resource estimation, such as WAsP, and supporters of statistical methods like MCP. According to the authors, the study shows that both WAsP and MCP methods are able to make useful forecasts of long-term wind conditions at the selected sites of the target area. They also state that WAsP functions quite well, despite violation of the assumptions underlying the flow model. Bowen and Mortensen [34], however, in a study they performed on WAsP prediction errors as a result of site orography, conclude that these errors may be significant if they exceed the performance envelopes of WAsP for the climate and terrain. In fact, according to Brower [9], WAsP failed when it was used for the first time in coastal mountain passes in the US (where the first wind farms of that country were constructed), as it was unable to accurately predict the wind resource distribution. Wind flow modelling was held in low esteem in US wind resource assessment circles for many years as a result of this experience. Nevertheless, Brower [9] admits that WAsP remains popular despite its known limitations. He considers this to be partly due to the fact that many wind project sites do not have steep terrain or significant mesoscale circulations.

Prasad and Bansal [16], using data compiled from various sources, constructed a table showing wind speed uncertainty depending on the approach used to estimate annual mean wind speed and mean wind speed uncertainty depending on the duration of the wind data series for the target site. According to the data shown in the table, the minimum wind speed uncertainty is obtained with the use of WAsP. However, the authors consider [16] that each of the different wind speed forecast methods will have its pros and cons at a particular site and it will be up to the planner to take a decision as to which method is the most suitable in each case. They point out, for example, that WAsP can display a high degree of uncertainty in wind speed prediction when dealing with a site of complex topography. So, MCP methods, which do not require topographic site details, would appear to be more suitable for sites of complex topography. They also make clear that the uncertainty of MCP methods will be lower if the duration of the wind speed measurement campaign at the target site is extended. In Refs. [35–36], the authors analyse that the use of WAshwhen obstacles are found near the anemometer mast.

MCP methods have been mainly applied in estimation of onshore wind resources [37–39], though their use has also been proposed for offshore wind resource estimation [40–43] and the possibility of their use has been analysed for swell estimation [40], given the dependence of swell on wind speed [44]. It can be seen that MCP methodology has spread to other renewable energy sources. In this context, studies have been undertaken with the

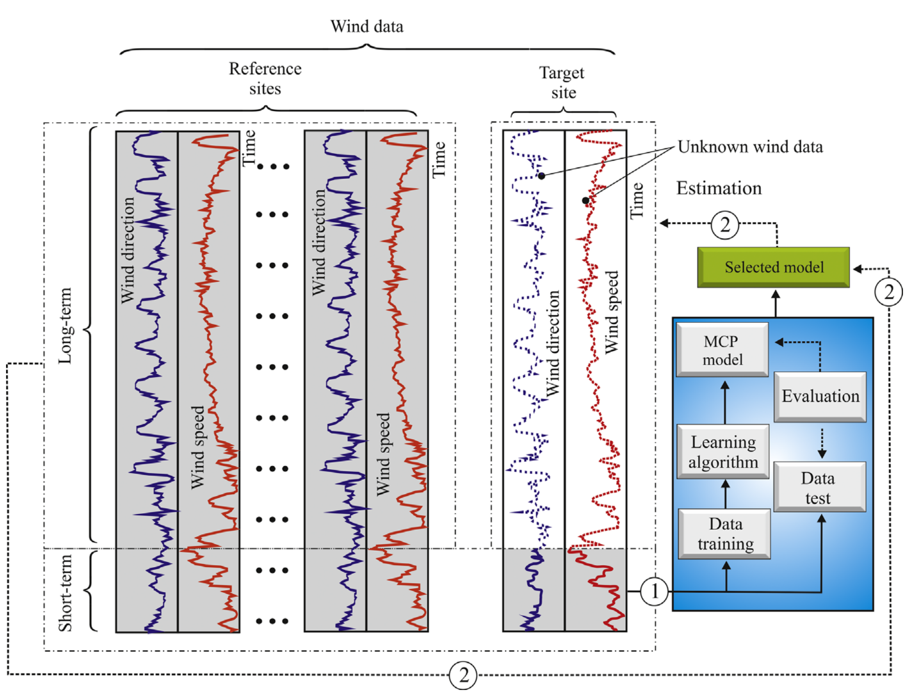

Fig. 1. Block diagram of the procedure normally employed by MCP methods.

aim of examining the quality and consistency of some MCP methods in the estimation of long-term solar resources (horizontal global radiation) at a target site [45].

Due to the characteristics of the variables that are involved in MCP methods, it is assumed that predictions obtained with them are subject to uncertainty [15,46–53]. With this in mind and given that the users do not have at their disposal long-term data for each site, several authors have proposed the use of statistical models to estimate this uncertainty [46–48,52]. These models attempt to estimate MCP prediction uncertainty based on the relationship established between data recorded simultaneously at the reference and target sites. This involves assuming that the variability of the relationship between the data from the reference and target sites, as seen in the concurrent data set, reflects the long-term variability between the data of the two sites.

1.3. Aim of this review

A review is undertaken in this paper for a wide range of MCP methods that have been proposed for wind energy analysis, some of which have been implemented in wind energy industry software [26–32] and analysed in academic environments [54–57]. This review includes the initial methods proposed in the 1940 s which generally only attempted to estimate the long-term annual mean wind speed from a single reference station, and extends to methods proposed in the present century based on automatic learning techniques which use several reference stations and which are not as yet in common use in the wind industry.

In addition to describing the linear, non-linear and probabilistic transfer functions used by the different algorithms, the hypotheses on which these algorithms are based and the data format with which they work (time series or frequency distributions), this review will also examine the limitations in the use of MCP methods, the uncertainty associated with them and the different reference data sources that have been proposed. In this sense, the extensive collection of MCP methods which have been brought together and reviewed in this paper, ranging from the simplest and easiest-to-use models to the most complicated computational ones which require specific user experience, comprises an extremely useful catalogue.

2. Determinants of MCP methods

It is important to note that MCP methods are based on a series of hypotheses. These hypotheses are described and analysed in the following sub-sections:

2.1. Appropriate measurement protocols

MCP methods are based on the hypothesis that the long-term data of the reference and target stations and the short-term data of the target station have been obtained in accordance with appropriate measurement protocols that can guarantee their accuracy [5,7,9]. Albers and Klug [7] report that, because of a lack of experience, many wind speed measurements have an unacceptably high uncertainty as proper practices were not observed when selecting and mounting anemometers, when choosing the measuring site, or when deciding upon the height and duration of the measurements. Before using wind data series in an MCP analysis, it is of paramount importance that a thorough check is made on their validity [5].

In order to pass the consistency test that this hypothesis entails, the reference stations must satisfy a series of requirements throughout the historical (long-term) reference period [11,58–59]:

(a) The weather station data must not have been affected by factors such as the construction of buildings, installation of wind farms [60] or changes to the vegetation surrounding the wind masts that might distort the relationship between the target and reference site wind data.
(b) Neither the height nor location of the wind mast can have been subject to modification during the period.
(c) Probst and Cárdenas [61] propose that the heights above ground level (agl) at which the data are collected at the reference and target sites should be similar. In a typical case in which the reference station data are recorded at a low height (generally 10 m agl) and the target site station data at a range of heights between 40 and 60 m, according to the authors, it has been shown that the correlation coefficient between the two data series is lower than that obtained when both data series are recorded at the same height of 10 m agl. They state that this is because there can be substantial variation between daily wind speed profiles at very different heights. The authors propose downscaling the wind speeds at the target site to 10 m agl before calculating the parameters of a linear regression based MCP method. Once the MCP has been applied and the long-term wind speeds estimated at the target site, these can then be rescaled to the corresponding height of 40 or 60 m. The authors [61] state that in this procedure case-by-case verification should be made that the errors generated in the extrapolation stage do not eliminate the reduction in the prediction error obtained as a result of the downscaling. With regard to this point, the authors of this review would like to underline this recommendation of Probst and Cárdenas concerning case-by-case verification, as in some cross-correlation analyses conducted between anemometer stations installed in the Canary Islands (Spain) results were obtained

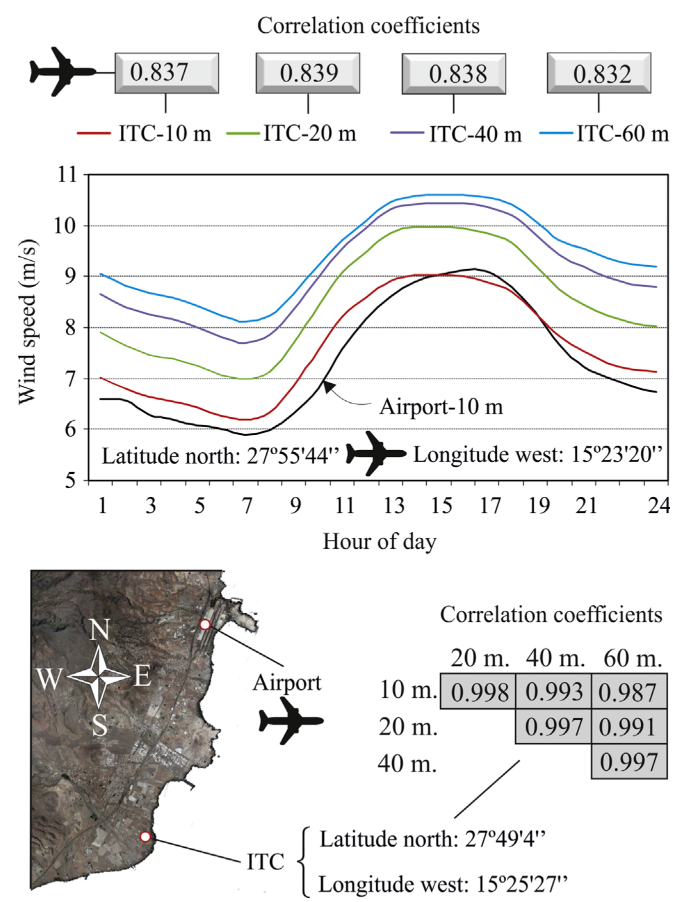

Fig. 2. Daily wind speed profiles, at different heights above ground level, from two anemometer stations installed on the island of Gran Canaria (Spain). Correlation coefficients between the wind speeds recorded in 2010 at the various heights.

which do not confirm the wisdom of applying the downscaling strategy proposed in [61]. For example, it can be observed in Fig. 2 how the cross-correlations made between hourly mean wind speeds recorded in 2010 at a target site (installations belonging to the Canary Islands Technological Institute) and reference site (Gran Canaria airport), some 13 km apart, display certain discrepancies with the results indicated in [61]. There are no significant differences between the correlation coefficients of the reference station data measured at 10 m agl and the data of the target site station measured at different heights (10, 20, 40 and 60 m agl). The highest correlation coefficient between the two stations is obtained between the data recorded at 10 m agl (reference) and 20 m agl (target). However, the correlation coefficients obtained between the data recorded at the target site station at different heights (10, 20, 40 and 60 m agl) reflected what was expected. That is, as also observed by Jain [2], the lowest correlation was found between the measurements at 60 m and 10 m agl (r=0.987) as a result of surface roughness of the ground. In addition, although ideally a measure of atmospheric stability (stable, neutrally stable or unstable) could be used, sufficient information may not be available to allow its calculation and in many situations the vertical wind profile models (logarithmic law and power law) usually only consider neutral stratification [2,9,16]. For this reason, Oliver and Zarling [62] propose a procedure to replace the effect of atmospheric stability, which consists of calculating the correlation according to the hour of the day, given the influence of daily changes in atmospheric thermal stratification on vertical wind profiles [63].

(d) The wind data must been recorded with essentially the same equipment.

2.2. Similar wind climate at the reference and target sites

This assumption is based on the idea that the reference station data are representative of the wind climate of the target site. Basically, this is usually equated in the literature to the existence of a good correlation between the reference station and target site station. However, other types of verification are also commonly performed. These include comparisons between daily and monthly wind speed profiles and comparisons between the wind roses recorded at the reference and target sites. Appreciable differences in these comparisons or low coefficients of correlation can lead to rejection of the reference station for its use in MCP methodologies [2,64–65].

The correlation is often quantified by calculating the Pearson product-moment coefficient of linear correlation (otherwise known as Pearson's correlation coefficient), r. In this case, Pearson's correlation coefficient measures the strength and direction (rising or falling, depending on whether the sign is positive or negative, respectively) of the linear association between the wind speeds recorded at the reference station and target site station during the concurrent data period [2,5,9]. Its value is given by the following equation [66–69]:

$$r = \frac{\sum_{i=1}^{n} [(v_i)_t^{ST} - \overline{v}_t^{ST}][(v_i)_r^{ST} - \overline{v}_r^{ST}]}{\left\{\sum_{i=1}^{n} [(v_i)_t^{ST} - \overline{v}_t^{ST}]^2\right\}^{1/2} \left\{\sum_{i=1}^{n} [(v_i)_r^{ST} - \overline{v}_r^{ST}]^2\right\}^{1/2}} \quad -1 \leq r \leq 1 \qquad (1)$$

where $(v_i)_r^{ST}$ and $(v_i)_t^{ST}$ are the observed short-term wind speeds at the reference and target sites, respectively. $\overline{v}_r^{ST}$ and $\overline{v}_t^{ST}$ are the respective means.

According to a number of references [2,5,9], if the correlation is weak the results can be misleading and, in such a case, the long-term data from the reference station should be discarded or used with precaution.

Clarification should be made at this point that the use of Pearson's correlation coefficient r to measure the degree of association between the wind speeds of the target and reference site can be a valuable tool. However, certain precautions need to be taken, as r is neither robust nor resistant [69]. It is not robust because strong but non-linear relationships between wind speeds at the two sites may not be recognised, and it is not resistant because it can be extremely sensitive to one or a few pairs of outliers. In other words, the value of r can be seriously affected by the presence of a few odd observations which are found far from the bulk of the data, Fig. 3. These 'odd' observations can generate a high asymmetry in the distributions of the two variables, and this asymmetry has a profound effect on the correlation coefficient.

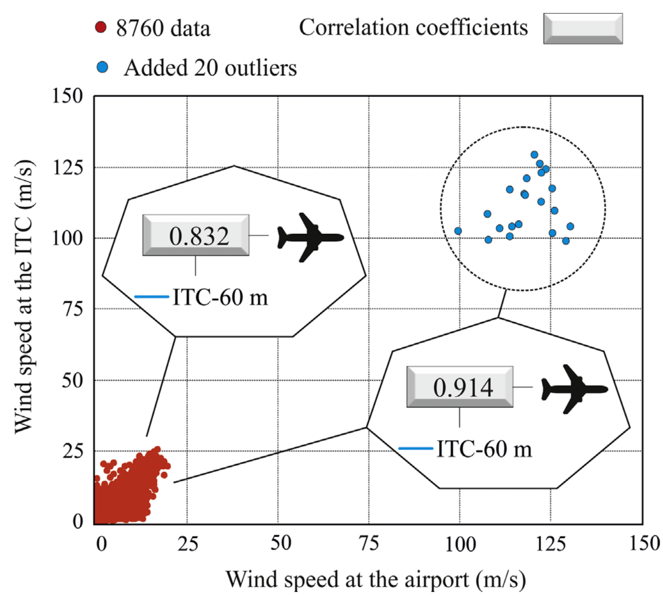

Fig. 3. Effect on the correlation coefficient of the data recorded during 2010 by two anemometer stations installed on the island of Gran Canaria (Spain) at 10 m agl (Airport) and 60 m agl (ITC), when outliers are introduced which produce asymmetry in both variables.

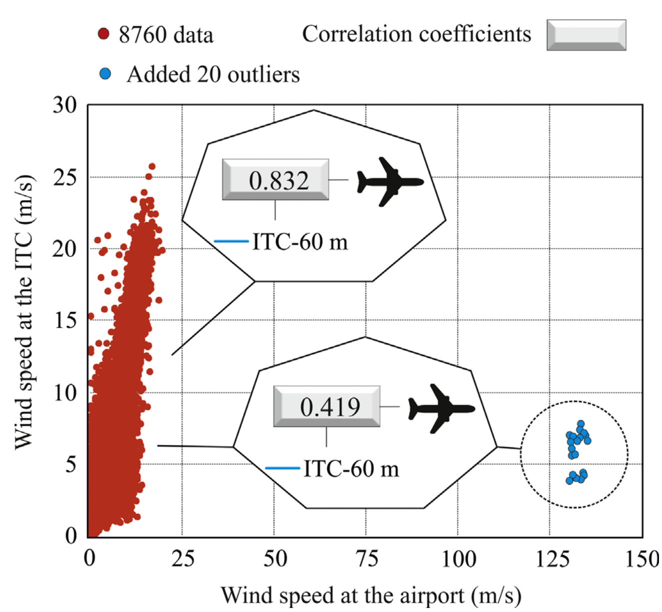

Fig. 4. Effect on the correlation coefficient of the data recorded during 2010 at two anemometer stations installed on the island of Gran Canaria (Spain) at 10 m agl (Airport) and 60 m agl (ITC), when outliers are introduced which produce asymmetry in just one of the variables.

In this case, the impact produced on r by the 'odd' observations is commonly called the "King-Kong" or "Big Apple" effect [67–68]. This effect can substantially increase the value of r. Sometimes an outlier can produce an asymmetric distribution in just one of the variables, lowering the value of r [67], Fig. 4. Finally, it should be noted that the value of r only indicates the extent to which the target and reference wind speeds are linearly associated. r in itself does not explain any relationship between the wind speeds at these sites, at least not in any physical or causative sense[4] [66,67].

McKenzie et al. [70] argue that not all MCP methodologies have a linear nature and it is therefore possible that ways of measuring the correlation other than Pearson's may provide a more precise description of the relationship between the wind data of the two sites. They claim, for example that a more appropriate method would be to use a non-parametric measure of association such as Spearman's rank correlation coefficient, r_s, [71]. This coefficient can be determined through Eq. (1), replacing the observations of the wind speeds recorded at the two sites with the ranks. Whereas Pearson's correlation coefficient is a measure of the degree of linear association that exists between the wind speeds recorded at the target and reference sites, Spearman's rank correlation coefficient measures the trend. In a sense, r_s has higher significance than r because, when measuring the degree of monotonic association between the wind speeds, r_s is not restricted to finding a linear association between them [71].

Several studies have been made in relation to the influence on the correlation coefficient of the time scale of the wind data series and the spatial distances between the reference and target anemometer stations [33,34,72,73]. Nielsen et al. [74] state that before using an MCP method consideration should be given to the possible need for time averaging of the data. If short time averages (approximately 10 min) are used but the stations are not closely situated, probably no correlation will be found between the wind speeds as a result of variations in the short time scale. Ayotte et al. [72] undertook a simple time and spatial analysis of wind speed and direction data obtained from a number of weather towers separated by distances ranging from 1 to 100 km. For this purpose, they developed a band-limited correlation coefficient in the frequency domain, performing discrete Fourier transforms [75] of the time series. Ayotte et al. [72] note that there is a close relationship between the distance that separates the two sites where the wind data were recorded and the correlation that exists between the data. This in turn means there will be a relationship between that distance and the accuracy that is expected when predicting the averaged wind speed and direction values based on these recorded data.

Bowen and Mortensen [34] obtained interesting results in their study, though these were not conclusive. Bearing in mind the significant distance between most of the sites analysed, they concluded that cross-correlation coefficients with zero time lag, calculated with 10-min averaged wind speed series, might not be suitable, deciding that a longer average time (specifically 1 h) might be more appropriate in these cases. However, they only observed a small improvement in the cross-correlation coefficients, Eq. (2), with a time lag k equal to zero for 1-hour averaged wind speeds, compared to the coefficients obtained with 10-min averaged speeds. The authors confess that they did not look into this question in great depth and did not resolve it:

$$r(k) = \frac{c(k)}{s_t^{ST} s_r^{ST}} \tag{2}$$

In the cross-correlation equation, $c(k)$ is given bythe following equation:

$$c(k) = \frac{1}{n} \sum_{tt=1}^{n-k} [(v_{tt})_t^{ST} - \overline{v}_t^{ST}][(v_{tt+k})_t^{ST} - \overline{v}_r^{ST}] \tag{3}$$

where $(v_{tt})_t^{ST}$ and $(v_{tt})_r^{ST}$ are, respectively, the short-term target and reference site wind speeds at time tt, s_t^{ST} and s_r^{ST} are their standard deviations and $\overline{v}_t^{ST}$ and $\overline{v}_r^{ST}$ are the mean values, $k(=0, 1, 2, 3,\ldots \mathrm{Kr})$ is the time lag, n is the number of data and Kr is the number of lags considered.

Früh [73] calculated the cross-correlations between hourly mean wind speeds of seven anemometer stations in Scotland. According to the author, the results were as expected. Good correlation coefficients were observed between $r=0.67$ and $r=0.84$ with 0 or 1 h time lag. The correlation coefficient between the westernmost and easternmost sites fell to 0.52 with a 3 h time lag. Früh [73] observed that the range of time lags for the best correlation between two sites was consistent with the approximately south-westerly prevailing wind. Lamberg and Mortensen [33], understanding that for proper use of MCP methods a good correlation between target and reference wind data is vital, also investigated this question in a study of 15 possible cross-correlations between 10-min averaged wind speeds from six sites. As all the correlation functions in their studies decreased, they concluded that it was not necessary to take into consideration any time lag.

No definitive and corroborated scientific criterion has been published in renewable energy related literature concerning the correlation level below which reference station data should not be used due to their unreliability [2,5]. However, some rules of thumb have been proposed [2,5,19,58,65,76]. According to Ref. [19], the coefficient of determination [75], R^2, should not be lower than 70%. R^2 is defined in general terms as the ratio between the variability explained by the regression and the total variability, as seen in the following equation:

$$R^2 = \frac{\sum_{i=1}^{n} [(\hat{v}_i)_t^{ST} - \overline{v}_t^{ST}]^2}{\sum_{i=1}^{n} [(v_i)_t^{ST} - \overline{v}_t^{ST}]^2} \tag{4}$$

R^2 measures the proportion of the total variation about the mean, $(\overline{v})_t^{ST}$, explained by the regression. Often, R^2 is expressed as a percentage and is the square of what is usually called the multiple correlation coefficient. In fact, R is the correlation between the measured short-term data, $(v_i)_t^{ST}$, and the short-term data as estimated by the model, $(\hat{v}_i)_t^{ST}$.

In Ref. [5], it is reported that, as a general rule, when monthly and in all directions the square of Pearson's correlation coefficient,[5] r^2, of the wind speed is less than 0.8 there is substantial uncertainty as to whether the reference station long-term data should be used to estimate long-term wind conditions at the target site.

According to Bass [76], care should generally be taken that the coefficient of correlation is higher than 0.7 and according to Jain [2] if the value of the coefficient of correlation is 0.9 or higher the correlation is considered to be excellent. A guide is also provided in Ref. [2] for the acceptable correlations when it comes to determining whether two wind speed time series share the same wind climate. So, for example, a correlation of raw data between wind speed series with 10-min intervals above 0.65 or a correlation of daily average data of those series above 0.75 is considered acceptable. In the study undertaken in [65], the sites which were considered clearly unsuitable for an MCP method were discarded. So, the appropriateness or inappropriateness of using an MCP

[4] Despite r being neither robust nor resistant it has been widely used, probably because its form lends itself well to mathematical manipulation and because it is closely related to the linear regression used in MCP methods.

[5] Only in the case of a simple linear regression is the coefficient R between $(v_i)_t^{ST}$ and $(\hat{v}_i)_t^{ST}$ exactly the same as the correlation coefficient r between $(v_i)_t^{ST}$ and $(v_i)_r^{ST}$. In other cases, the relationship $R^2 = r^2$ does not hold.

method is based on the level of correlation. In this context, overall correlation coefficients[6] of 0.5–0.6 are considered to be very poor, and those with ranges of 0.6–0.7, 0.7–0.8, 0.8–0.9 and 0.9–1.0 are considered, respectively, poor, moderate, good and very good. These ratings which are dependent on the correlation coefficient value are also set out in Ref. [58].

2.3. Knowledge of the pattern of seasonal variations in the concurrent data period

The short-term period over which wind data is available for both the reference and target sites must be at least long enough to allow information to be extracted with regards to seasonal wind variations (speed and direction). The general recommendation is for this concurrent data period to be at least one year long. Taylor et al. [49] undertook a study to analyse the long-term prediction variability of MCP methods against annual data. For this purpose, they used sub-periods of 1, 3, 6, 12, 18 and 24 months. Their results suggest that one month of data gives rise to a margin of uncertainty (relative standard deviation) of between 6.5% and about 12% for long-term wind speed estimations at the analysed stations. For the three and six month periods, in spite of the existence of an inverse relationship between the uncertainty and the data measurement period, the difference between the margins of uncertainty is still quite high (approximately 4% for the 6-month sub-period). However, once the data period is extended to 12 months there is a considerable fall in uncertainty and the difference between the margins of uncertainty is significantly smaller (approximately 2%). Few changes are noted in uncertainties in the analysis of the 18-month sub-period, but the 24-month period sees a reappearance of the trend towards lower uncertainty. The authors conclude that the results obtained from their study provide evidence that there is a significant seasonal influence on the quality of long-term estimations. Consequently, they underline the importance of having at least 12 months' (or multiples thereof) worth of measurements for the target site. Oliver and Zarling [77] state that long-term wind speed predictions made with less than one year's worth of data for the site can be subject to large errors. The clear seasonal patterns that appeared in a study carried out by these authors using 14 station pairs (target–reference) situated in the "wind corridor" from Texas to Dakotas, showed the importance of taking into consideration exactly when during a year a measurement campaign is initiated when making a prediction of wind speed for data sets of less than one year's duration. Often, business considerations [4,15,77] will prevent a year's wait for data records and will require predictions to be made on the basis of less than one year's data for the target site. In these cases, a greater understanding is required for the effects of seasonal fluctuations on the bias of the wind speed prediction so that a more informed analysis can be made of the long-term wind speed prediction [77].

2.4. Climate stability

MCP methods work on the assumption that the data sets are statistically stationary. That is, the effects of climate change are ignored and it is assumed that wind behaviour in the future, during the life cycle of the energy project, will be similar to that of the past. Wind and climate variabilities are closely linked. In scientific circles, it is generally reckoned that as a result of global warming the difference in temperature between the poles and the equator will be reduced and mid-latitude winds will also undergo some modifications [63,78]. Nonetheless, it should be pointed out that in the specialist literature consulted by the authors of this review there exists a certain amount of controversy as to the variation experienced in the last few decades by wind resources throughout the world as a result of climate change or alterations to surface roughness of the ground. A number of researchers have reported that surface winds have fallen in China, the Netherlands, the Czech Republic, the United States and Australia in the last few decades [79–80]. Vautard et al. [81] analysed the extent and possible cause of surface wind speed changes in the northern mid-latitudes between 1979 and 2008 using data from 822 surface weather stations. They reported that surface wind speeds declined between 5% and 15% in almost all continental areas in the northern mid-latitudes, and that strong winds slowed faster than weak winds. The mesoscale model simulations that they used suggest that an increase in surface roughness could explain reduction of the wind between 25% and 60%. However, other authors [9,82] claim that there is thus far little reliable evidence of a significant rise or fall in wind resources in the world over the past few decades as a result of climate change. Several studies have been carried out with the aim of analysing potential changes in wind speed in the future as a result of climate change [77–87]. Breslow and Sailor [83] investigated potential impacts throughout the continental US. For this purpose, they used two general circulation models which were, in general, consistent in the prediction that the US will see a reduction compared with the 1948–1975 period in the mean annual wind speed of between 1.0% and 3.2% in the next 50 years and between 1.4% and 4.5% in the next 100 years, as a consequence of the response of the atmosphere to climate change induced by the increase in CO_2 concentration. On a regional scale, the two models showed some similarities in the first years of simulations (2050), but diverged significantly in their predictions for 2100. So, there is clearly a lot of uncertainty as to how wind fields will change in the future [84]. Sailor et al. [85] used statistically 'downscaled' outputs of four general circulation models (GCMs) to investigate scenarios of the impact of climate change on wind energy generation in five states in the northwest of the United States. The results suggest that summer wind speeds in the northwest may fall by 5–10%, whereas winter wind speeds will suffer only a moderate decline or may even rise slightly. Watson and Kritharas [86] used historical variability (53 years) of wind speed in the United Kingdom in order to evaluate the uncertainties in long-term assessment of the wind resource. Their analysis of long-term variability was carried out through the construction of a wind index based on 53 years of wind speed data from seven surface stations and 27 years of wind speed data from 57 stations in the UK. This index was also used to construct six regional indices. The authors concluded that the long-term data of 53 years do not appear to suggest any perceptible trend towards an increase or decrease in wind speed with time. The trends observed in the regional indices were not incompatible with the predictions of some regional climate models, but long-term trends for the prediction of future wind speeds remain very uncertain and there exists significant interannual variation in the observed data.

Pryor and Barthelmie [87] report that current state-of-the-art does not suggest the existence of detectable changes to the wind resource or other external conditions that could put at risk continued wind energy exploitation in the north of Europe, though more research is required to ensure a greater degree of confidence in such projections. It can be concluded from the results of the various studies that have been carried out that, as indicated by Brower [9], the effect of climate change on wind speed within the time horizon of wind project investments (i.e. up to 25 years) will probably be low.

3. Proposed reference data sources

Given the existence of sites where the restrictions listed in Section 2 limit the surface weather station reference data and the length of the reference periods that can be used, other data

[6] These come from the correlation calculated for each wind direction sector weighted with the frequency of this distribution sector at the reference site.

sources have been analysed [59,88,89] and studies undertaken to find out whether the use of reanalysis data might constitute an alternative approach [9,90–92]. One alternative source for reference data are data from rawinsonde measurements [59,88]. However, despite the consistency and reliability of these sources, there are considerably fewer of this type of station compared to surface stations and, generally, they collect data only on a 12 hourly basis at heights above the planetary boundary layer. Consequently, rawinsonde data do not correlate as well as surface station data with measurements taken at wind project sites [89].

Reanalysis is a method used to construct a record of climate data in a three-dimensional global or regional grid which combines past observations from different observation and measurement systems (surface weather stations, satellites, rawinsonde, etc.) with a numerical weather prediction (NWP) model that creates a physically consistent picture of how the earth's climate has evolved over time.

The most well-known reanalysis data are generated by the National Centres for Environmental Prediction (NCEP) [93] and the National Centre for Atmospheric Research (NCAR) [93] in the US and the European Centre for Medium-Range Weather Forecasts (ECMWF) [94] in Europe.

According to Brower [90], there are two main reasons that suggest that reanalysis data might be attractive for use in MCP methods. First is their ease-of-use, since the data are open for public use in a global network and there is no need to search for suitable weather stations or compare several different stations in order to find the best one. The second reason is that the data record is longer than that available for most weather stations.

However, Brower [90], as the main conclusion of studies undertaken with reanalysis data and direct observations, reports that NCAR/NCEP reanalysis data are not sufficiently reliable for use in MCP methods. Though the data reference periods are relatively short (10 years or less), the false fluctuations and trends that can occur in reanalysis data often result in bigger errors than those obtained with the conventional method. So, it is concluded [59,90] that although these errors do not result in all cases, bearing in mind the importance of MCP in energy production estimation prior to the construction of a wind farm, NCAR/NCEP reanalysis data should not be used for this purpose.

Liléo and Petrik [91] investigated the applicability of wind data from three different reanalysis data sets as long-term references in MCP analyses. In addition to NCEP/NCAR data, which have been commonly used in wind resource analyses over the past decade, they also used in their study MERRA (Modern Era Retrospective-analysis for Research and Applications) [95] and NCEP/CFSR (National Centres for Environmental Prediction /Climate Forecast System Reanalysis) data [96]. In their study they only analysed grid data covering the territory of Sweden.

The results obtained by the authors [91] suggest that the use of MERRA and NCEP/CFSR reanalysis data sets represents a relevant improvement in the accuracy of the MCP results. The higher spatial and temporal resolutions of the reanalysis data sets in comparison with the NCEP/NCAR data allow better representation of the local wind climate, as is demonstrated, according to the authors, by the high levels of correlation with the data records obtained with the anemometers installed on local masts.

Pinto et al. [92] compared NCEP/NCAR reanalysis data with local measurements from 20 wind measuring stations spread out over Portuguese territory. Among the different conclusions made in the study, they report that the 6 h data from the reanalysis project displayed a poor correlation with the measurements of the local stations as a result of the high variability in both data series. However, they indicate that the correlation tended to improves when the data were compared on a monthly and annual basis.

Taylor et al. [89] presented windTrends, a database of meteorological conditions covering the years 1997–2009. WindTrends offers a snapshot of the climate each hour at different heights agl in a regular-sized grid with 20-km resolution (around 100,000 data records are available for each height and 20-km point). The authors provide a summary of the results they obtained in a thorough evaluation of the suitability of the data for use in MCP methods and other applications. Though conceptually similar to reanalysis, the process differs in two fundamental aspects: grid resolution is finer and the observational data source – rawinsonde only – remains the same throughout the simulation period. According to the authors, the validation study they presented shows that windTrends is generally superior to global reanalysis [97] and comparable with ASOS (Automated Surface Observing System) stations for the purpose of use in MCP methods.

4. Overview of proposed measure-correlate-predict models

4.1. Considerations about wind speeds used

It should be noted that most MCP methods, for the purpose of improving predictability, do not use the full range of available wind data [2,5,33,65,98–99]. Filtering of the wind speed data is usually proposed along with rejection of data that fall below a certain threshold value [2,5,33,65,98] or different treatments of values below this threshold [98,100]. According to some authors

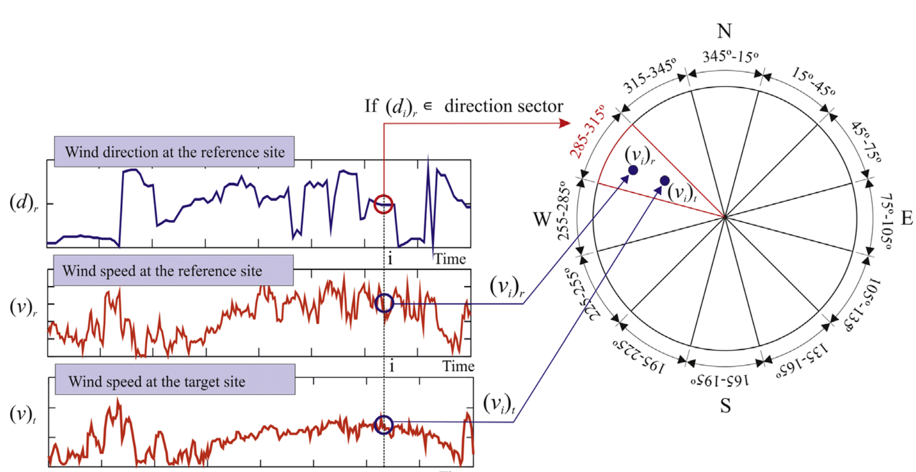

Fig. 5. Wind direction sectors of equal size, commonly used in MCP methods.

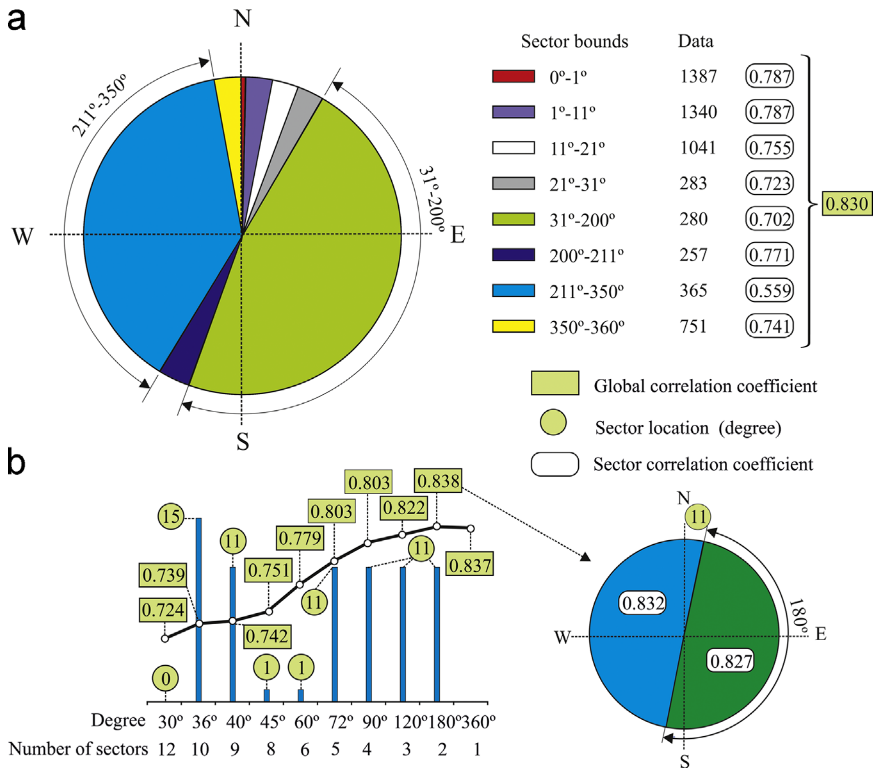

Fig. 6. Results of applying two methods of determining the number and size of the wind direction sectors which avoid the existence of sectors with an insufficient number of data to the hourly mean data recorded at 10 m agl during 2010 at two anemometer stations (Airport and ITC) installed on the island of Gran Canaria (Spain). (a) Method which aims for each sector to have at least 1/24 of the total number of data, (b) Method which aims to obtain the maximum overall coefficient correlation.

[2,5,33], wind speeds below 3 m/s are not important from an energy point of view and can skew the functional relationship. However, in the analysis undertaken in [65] calms (wind speeds below 0.1 m/s) constituted this threshold value.

4.2. Considerations about the influence of wind direction

Generally, in the most widely used MCP methods in the wind industry or proposed in the scientific literature, the relationships established between the target and reference stations are not constructed using the whole mass of data together. Instead, to take into consideration the effects of wind direction, the data are binned or grouped in sectors defined by the reference site wind direction and a relationship is then constructed for each different sector. The wind directions recorded at the reference and target sites are compared to determine whether there exist local characteristics which might have an influence on the directional results [5]. If the differences between the wind direction records at the target and reference sites are substantial various approaches may be taken [61]. Woods and Watson [101], Velázquez et al. [102] and Carta et al. [103] have also proposed other approaches which are described in Sections 4.4.2 and 4.5.

The direction sectors are usually of equal size, with 12 sectors of 30° commonly being employed [5,98,104–106], Fig. 5. However, care should be taken when defining the number of direction sectors, as it may happen that there are sectors for which an insufficient number of data are available to provide a reliable fit. To reduce the effect of this problem, Riedel et al. [107] and King and Hurley [98] propose that the direction sectors should be determined dynamically, in such a way that each sector contains an equal number of data or an equal fraction of the total data. More specifically, Riedel et al. [107], in studies carried out with hourly and 10-min mean wind speeds, report that they obtained good results when each sector contained at least 1/24 of the total data.

A method called "RES-OptiSector" is described in a final report [104] which describes the work undertaken in a research project entitled: "An improved measure-correlate predict algorithm for the prediction of the long term wind climate in regions of complex environment", funded in part by the European Commission within the framework of the Non-Nuclear Energy Programme JOULE III (Programme Acronym: FP4-NNE-JOULE C) [108] (contract number JOR3-CT98-0295). This method automatically identifies the optimum number of direction sectors and their location with the aim of maximising the overall[7] Pearson's correlation coefficient, r,[8] between the concurrent wind speed data of the reference and target sites. For this purpose, the researchers tested various optimisation algorithms and the most robust was found to be one based on a Monte Carlo simulation. The algorithm starts by calculating the value of r for just one sector. It then determines and stores the optimum value of r of the set of tests generated with two sectors, repeating this process by adding one sector until a total of 12 has been reached, Fig. 6. In [104], a method called "RES Fourier" is described which attempts to avoid the problem of a wind direction sector containing just a few data points. For this, the use was proposed, as an alternative to the classical linear regression, $f[\cdot]$ (with slope β and offset α), of the function shown in the following equation:

$$(\hat{v}_j)_t^{ST} = \left\{ a_0 + \sum_{k=1}^{M} a_k \sin[(\theta_i)_r^{ST}] + b_k \cos[(\theta_i)_r^{ST}] \right\} f[(v_j)_r^{ST}, (\theta_i)_r^{ST}] \qquad (5)$$

The number of considered equal direction sectors, $(\theta_i)_r^{ST}$, was i=36 and the number of independent parameters $[2M(a_k, b_k) + 1(a_0) + 2(\alpha, \beta)]$ used was 19. These parameters were determined by

[7] These come from the correlation calculated for each wind direction sector weighted with the frequency of this distribution sector at the reference site.

[8] Eq. (1) is used, but the variables refer to each direction sector.

minimising the chi-square statistic with the simplex method [75]. As reported in [104], the number of independent parameters was obtained after reaching a reasonable compromise between accuracy and computational speed.

Oliver and Zarling [62] propose the binning or grouping of data by time of day as a way of considering the effect of atmospheric stability. These authors [62] analysed 19 pairs of target-reference sites with at least two years' worth of concurrent data. They grouped the data into 12 divisions of the day (each division comprising 2 h) and compared the results with the traditional approach of grouping into 12 wind direction sectors and with combinations of both approaches.

They concluded that, in most cases, choosing a combination of six direction sectors (60° each sector) and two time groups (12 h each group), instead of the traditional approach of 12 direction sectors (30° each sector), gave a better prediction. It was postulated that by maintaining the traditional 12 direction groups and adding 2 or 3 time-based groups more the predictions will be improved while preserving sufficient data points in each group. It was also considered that some further improvement may be achieved by varying the time group widths depending on the month of the year.

4.3. Estimation methods of the long-term mean wind speed

The predictions methods of long-term wind characteristics which were initially proposed between the 1940s and 1980s [10,64,109–121] generally try to estimate the mean annual wind speed using just one single reference station. In 1948, Putnam [109] proposed estimating the long-term mean wind speed, $\overline{v}_t^{LT}$, at the target site station based on the knowledge of its short-term mean wind speed, $\overline{v}_t^{ST}$, using a procedure known as climatological reduction by the method of ratios [10,103,109,113,115,116], if the site in question happens to be in the vicinity of a long-established meteorological station. The proposal of Putnam [109] was to use the following equation:

$$\overline{v}_t^{LT} = [\overline{v}_t^{ST}/\overline{v}_r^{ST}]\overline{v}_r^{LT} \tag{6}$$

In Eq. (6), $\overline{v}_r^{LT}$ is the long-term mean at the reference site and $\overline{v}_r^{ST}$ its short-term mean (for the period coinciding with that of $\overline{v}_t^{ST}$).

Conrad and Pollak [115] state that the requirements that need to be met for this method to be applicable are that the deviations from the climatic mean must be quasi-constant and relatively homogenous. The 'quasi-constant' requirement is mathematically equivalent to that of a high spatial cross-correlation between the deviations of the climatic means at the target and reference sites.

Corotis [111] proposed a method subsequently referenced in various studies [10,103,110,112,114–116] which explicitly includes long-term spatial cross-correlation, r^{LT}, between both sites and the long-term standard deviations at the target site, s_t^{LT}, and reference site, s_r^{LT}:

$$\overline{v}_t^{LT} = \overline{v}_t^{ST} + r^{LT}[\overline{v}_r^{LT} - \overline{v}_r^{ST}][s_t^{LT}/s_r^{LT}] \tag{7}$$

As the long-term cross-correlation and long-term standard deviation are unknown for the target site, Daniels and Schroeder [112] carried out a series of hypothesis tests to estimate them. For the hypothesis that there is little variation in the correlation over time, they proposed use of the short-term correlation instead of the long-term correlation. The long-term variance at the target site is estimated by the following equation:

$$(s_t^{LT})^2 = (s_t^{ST})^2 + (s_r^{LT-ST})^2 \tag{8}$$

The final term of Eq. (8) is the long term variance of reference site mean wind speeds averaged over periods as long as the short-term survey period, ST. This expression assumes that this variance is the same at the target and reference sites. In addition, it considers that the variance of the mean hourly wind speeds at the target site during the short-term period is representative of the long-term conditions.

Justus et al. [10] examined the procedures of Putnam [109] and Corotis [111] and published the accuracy of the results they obtained in climatological fits performed with wind data from various anemometer stations at different locations in the US. Only in two of the six pairs of cases analysed were the errors between the long-term mean wind speeds estimated with the methods of Putnam and Corotis and the observed wind speeds lower than the errors obtained between the observed long-term mean wind speeds and the short-term annual mean wind speed measured at the target sites.

Though both the method of ratios [109] and the method of Corotis [111] were initially used without taking into consideration wind direction, proposed variations do include consideration of the influence of wind direction. Methods have been proposed which differ from that of Putnam [109] in which different ratios are used, namely one ratio for each of the N sectors of $\theta°$ defined by the wind direction sector $(\theta_k)_r$ of the reference site. That is, Eqs. (6) and (7) are not applied to the whole mass of data together. Instead, in order to take into consideration the effects of wind direction, the wind data are binned or grouped in sectors defined by the wind direction at the reference site and these equations are then used for each of the different sectors.

Harstveit [118] proposes using the so-called KH method, in which a ratio of mean wind speeds is used for each wind direction sector. However, this procedure belongs to the small group of methods which do not follow the typical structure shown in Fig. 1, as each ratio determines the relationship between the short-term and long-term wind speeds at the reference station. The KH method, proposed by Harstveit [118] to predict the long-term wind speed in a particular wind direction sector,$(\theta_j)_t$, of the target site weights the contributions of each of the wind direction sectors, $(\theta_k)_r$, of the reference site:

$$[\overline{v}_t^{LT}]_{(\theta_j)_t} = \sum_{k=1}^{12}\left\{[\overline{v}_t^{ST}]_{(\theta_k)_r}\right\}\left\{\frac{[\overline{v}_r^{LT}]_{(\theta_k)_r}}{[\overline{v}_r^{ST}]_{(\theta_k)_r}}\right\}\left\{\frac{p^{ST}[(\theta_j)_t|(\theta_k)_r]}{p^{LT}[(\theta_j)_t]}p^{LT}\left[(\theta_k)_r\right]\right\} \tag{9}$$

The probability, $p^{LT}[\cdot]$, that the wind blows long-term in a direction sector j of the target site is estimated making use of the probability that the wind blows long-term in direction k of the reference site and making use of the observed conditional probability, $p^{ST}[\cdot|\cdot]$, at the target site during the short-term measurement period:

$$p^{LT}[(\theta_j)_t] = \sum_{k=1}^{12}\{p^{LT}[(\theta_k)_r]\}\{p^{ST}[(\theta_j)_t|(\theta_k)_r]\} \tag{10}$$

The Tallhaug and Nygaard method [117–119] is to a certain extent a variation of the method of Corotis [111]. Tallhaug's method estimates the long-term wind speed at the target site for each given wind direction sector k of the reference station through the following equation:

$$[\overline{v}_t^{LT}]_{(\theta_k)_r} = [\overline{v}_t^{ST}]_{(\theta_k)_r} + [r^{ST}]_{(\theta_k)_r}\{[\overline{v}_r^{LT}]_{(\theta_k)_r} - [\overline{v}_r^{ST}]_{(\theta_k)_r}\}\{[s_t^{ST}]_{(\theta_k)_r}/[s_r^{ST}]_{(\theta_k)_r}\} \tag{11}$$

The long-term wind speed in each wind direction sector j of the target site is estimated by the Tallhaug and Nygaard method [117–118] using the following equation:

$$[\overline{v}_t^{LT}]_{(\theta_j)_t} = \sum_{k=1}^{12}(\overline{v}_t^{LT})_{|(\theta_k)_r}\left\{p^{ST}\left[(\theta_j)_t\middle|(\theta_k)_r\right]\right\}\frac{p^{LT}[(\theta_k)_r]}{p^{LT}[(\theta_j)_t]} \tag{12}$$

The long-term directional probability at the target site is estimated using Eq. (10).

Tallhaug and Nygaard [117] do not recommend their model for mean wind speed estimation by wind direction sector unless the correlation is very high.

In 1982, Barchet [120], in a report for the Pacific Northwest Laboratory, described the use of weather pattern typing for characterising the wind regime of the Great Plains. The nine weather patterns that the author defined were based on the main features of fronts and air masses associated with a Polar Front Model of a synoptic-scale cyclonic storm system. In a later report, Barchet and David [113] characterised the wind regime of each weather pattern type (WPT) so that the long-term mean wind speed could be reconstructed from the mean wind speed associated with each of the nine WPTs and their frequency of occurrence. Barchet and David [113] used a reference station to help to calibrate the mean wind speed for a WPT and proposed the following equation to estimate the long-term mean wind speed at the target site. This extrapolation from short records is known as the "adjusted WPT" technique:

$$\bar{v}_t^{LT} = \sum_{j=1}^{9} (FO_j)_t (\bar{v}_r^{LT})_j \frac{(\bar{v}_t^{ST})_j}{(\bar{v}_r^{ST})_j} \qquad (13)$$

In Eq. (13), $(FO_j)_t$ is the frequency of occurrence of the WPT j at the target site [120–121]. $(v_t^{ST})_j$ and $(v_r^{ST})_j$ are the short-term mean observed wind speeds for the WPT j at the target and reference sites, respectively. $(v_r^{LT})_j$ is the long-term mean wind speed for the WPT j at the reference site. According to Barchet and David [113], a notable drawback of WPT techniques is that a quite sophisticated analysis needs to be performed of the meteorology of the area for them to be of use.

4.4. Estimation methods of the long-term wind characteristics with the support of a single reference station

These MCP methods were published fundamentally from the 1990s onwards. The most commonly employed methods use a single reference station and algorithms which are based on linear functions or models[9] [2,4,9,13,17,19,21,61,65,74,76,98–101,103–106,122–129] to establish the relationship between hourly (or a lower time interval) wind characteristics of the target site and the site chosen as the reference. Some of these methods can be used to estimate both wind direction and wind speed characteristics. The format of the data normally involves time series of frequency tables. The linear relationships between wind speeds proposed in renewable energy related literature include wind speed or frequency ratios [65,98,100,104,129], linear regression (with first-order linear models[10]) [2,5,9,13,17,19,61,65,74,76,99,101,103–106,123–125,127], the method of bins proposed by Beltrán et al. [99], the Vertical Slice method proposed by LeBlanc et al. [128], the SpeedSort, DynaSort and Scatter methods proposed by King and Hurley [98] and the variance method proposed by Rogers et al. [106] and subsequently referenced in various studies [4,21,61,125–128]. However, higher than first-order linear models have also been proposed as well as models which establish non-linear relationships [17,24,28,57,60,63,78,100,107–109,127,132–137] or probabilistic relationships [2,9,61,76,105,125–127,129–131,133,136–143].

4.4.1. Methods of ratios

The ratio of means with a single sector method [65] differs from the method proposed by Putnam [109] in that the ratio is applied to the reference site time series of long-term wind speeds

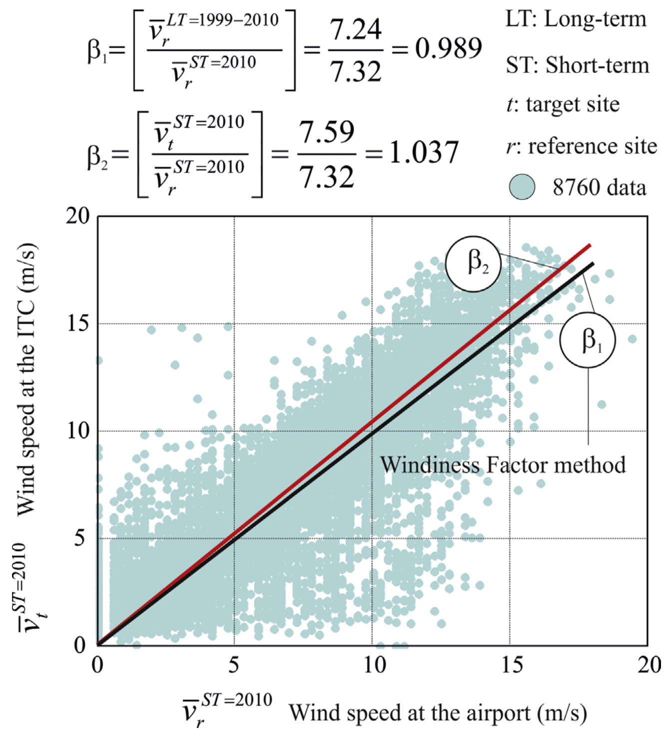

Fig. 7. Results of applying the single sector ratio method and the WF to the hourly mean data recorded at 10 m agl during 2010 at two anemometer stations (Airport and ITC) installed on the island of Gran Canaria (Spain).

instead of to the reference site's long-term mean wind speed, Eq. (14). This ratio is equal to the slope of a line which passes through the origin and centroid of the data, Fig. 6:

$$(v_j)_t^{LT} = \left[\frac{(\bar{v}_t^{ST})}{(\bar{v}_r^{ST})}\right](v_j)_r^{LT} \qquad (14)$$

The ratio of means with multiple sectors method [65] differs from the single sector version in that it tries to take into account the influence of the wind direction and for that purpose uses a ratio for each of the N sectors of $\theta°$ that can be defined in relation to the direction sector, $(\theta_k)_r$, of the reference site, Eq. (15). A total of 12 ratios are suggested in Ref. [104]. That is, a ratio for each of 12 direction sectors of 30° (345°–15°, 15°–45°, 45°–75°, 75°–105°, 105°–135°,135°–165°, 165°–195°, 195°–225°, 225°–255°, 255° –285°, 285°–315°, 315°–345°).

$$[(v_j)_t^{LT}]_{(\theta_k)_r} = \left\{\frac{[\bar{V}_t^{ST}]_{(\theta_k)_r}}{[\bar{V}_r^{ST}]_{(\theta_k)_r}}\right\}[(v_j)_r^{LT}]_{(\theta_k)_r} \qquad (15)$$

The EWEA method uses a ratio of means with multiple sectors and in its algorithm [98] the data of both sites with wind speeds below a certain cut-off speed are eliminated and the remaining data are distributed in 12 direction sectors defined by the reference site wind direction. A ratio of means is then defined for each sector and Eq. (15) is used to estimate the long-term wind speeds at the target site.[11] It is assumed that the long-term wind directions for the target site coincide with the long-term directions recorded at the reference site. This last assumption, according to King and Hurley [98], has led to poor sectorial accuracy in the studies. The authors point out they often observed that when the target site was windier than the reference site, the ratio between the target and reference sites' wind speeds fell as the

[9] In this paper, when we say a model is linear or non-linear we are referring to linearity or non-linearity in the parameters [66,71].

[10] The value of the highest power of the predictor variable in the model is known as the order of the model [66].

[11] If there are no data for a particular sector, the ratio is established at a predetermined value based on data obtained from all the sectors (that is, without taking into consideration the wind direction).

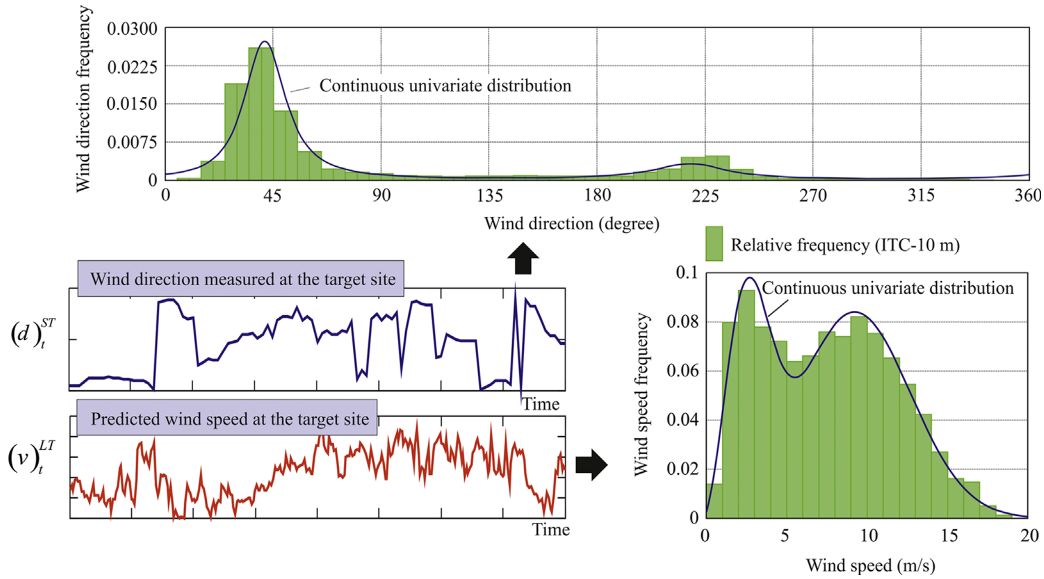

Fig. 8. Fit of continuous probability distributions to the estimated wind speed and direction histograms of the anemometer station located in Pozo Izquierdo (ITC, Gran Canaria-Spain), when using the WF method and long-term wind data (1999–2010) recorded at Gran Canaria airport.

wind speed increased. In such cases, the best relationship between them was provided by a curve rather than a straight line through the origin.

In [100], King and Hurley described two MCP methods which they call the Windiness Factor method (WF) and the Moulded Site Data (MSD) or WindScale method. In both methods, it is assumed that the long-term wind direction at the target site coincides with its recorded short-term wind direction.

In the WF method, which uses a single direction sector, a ratio is defined between the long-term mean wind speeds, $\overline{v}_r^{LT}$, and short-term mean wind speeds, $\overline{v}_r^{ST}$, at the reference site. This procedure, the analytic expression for which can be consulted in Eq. (16), belongs to the small group of methods which do not follow the typical structure shown in Fig. 1. In this case, the ratio or relationship is established between the data of the reference station and not between the short-term data of the target and reference stations. Once this ratio has been determined, adjustment of each of the short-term wind speeds recorded at the target site is performed by multiplying them by the windiness factor, Fig. 7. This set of modified or scaled wind speeds represents the long-term wind speeds at the target site:

$$(v_j)_t^{LT} = (v_j)_t^{ST}\left[\frac{\overline{v}_r^{LT}}{\overline{v}_r^{ST}}\right] \tag{16}$$

Using the time series of wind speeds generated with Eq. (16), together with the time series of wind directions recorded at the target site, the authors propose the construction of a table of relative frequencies, representing the long-term wind resource at the target site.

From the table of frequencies obtained with the WF method, it is possible to create wind speed and direction histograms to which continuous probability distributions can be fitted [144–145], Fig. 8. Likewise, a continuous probabilistic representation can be obtained of the long-term wind resource at the target site by fitting to the values of these tables a joint probability density function of wind speed and direction for subsequent wind energy analysis [146]. The MSD method [100] resembles the WF method in that the relationship is established between the short- and long-term data of the reference station and not between the short-term data of the target and reference stations, as occurs in most MCP methods. Initially, in the procedure followed by the MSD method, two table of frequencies are defined of wind speed and wind direction at the reference site, Fig. 9: one table of short-term frequencies, $(f_{i,j})_r^{ST}$, and another of long-term frequencies,$(f_{i,j})_r^{LT}$. For the construction of the two tables, *NB* wind speed bins and 13 wind direction sectors are defined. Twelve sectors of 30° (360°/12) and a thirteenth sector which will contain the wind speeds of the time series which are lower than a specified cut-off speed. That is, the thirteenth sector will contain all the wind speeds lower than the cut-off speed which would otherwise be in the other 12 direction sectors. King and Hurley [100] use the cut-off speed 3 m/s when $\overline{v}_r^{LT} < 6\ \mathrm{m/s}$, otherwise the cut-off speed is taken as 4 m/s.

In each of the two tables of frequencies and for each of the 13 direction sectors (13 columns of the frequency tables), the total frequencies are calculated, Eq. (17), as well as the mean wind speeds, Eq. (18):

$$(f_j)_r^{ST} = \sum_{i=1}^{i=NB}(f_{i,j})_r^{ST};\ (f_j)_r^{LT} = \sum_{i=1}^{i=NB}(f_{i,j})_r^{LT} \tag{17}$$

$$(\overline{v}_j)_r^{ST} = \frac{\sum_{i=1}^{NB}(f_{i,j})_r^{ST}b_i}{\sum_{i=1}^{NB}(f_{i,j})_r^{ST}};\ (\overline{v}_j)_r^{LT} = \frac{\sum_{i=1}^{NB}(f_{i,j})_r^{LT}b_i}{\sum_{i=1}^{NB}(f_{i,j})_r^{LT}} \tag{18}$$

Thirteen ratios of each parameter are determined, Eq. (19), from the values obtained with Eqs. (17) and (18). In order to lower the error produced when few data are available for a direction sector, if the frequency ratio is within the range from 0.2 to 5, its value is taken as 1. Likewise, the wind speed ratio is taken as 1 if it is within the range from 0.33 to 3. According to the authors, in the studies they carried out they found that the method was not sensitive to the exact specified range

$$Rv_j = \frac{(\overline{v}_j)_r^{LT}}{(\overline{v}_j)_r^{ST}};\ Rf_j = \frac{(f_j)_r^{LT}}{(f_j)_r^{ST}} \tag{19}$$

Each of the variables *k* of the short-term time series recorded at the target site is fitted by multiplying it by the appropriate wind speed ratio, Rv_j, and each time interval (for example, 1 h) is modified or scaled by multiplying it by the corresponding frequency ratio, Rf_j[12]:

$$(v_k)_t^{LT} = (v_k)_t^{ST}Rv_j;\ (t_k)_t^{LT} = (t_k)_t^{ST}Rf_j;\ (v_k)_t^{ST} \in \ \text{sector}\ j \tag{20}$$

[12] If there are sufficient data for all the sectors, for both the short- and long-term period, then the total time of adjusted hours will be equal to the total time

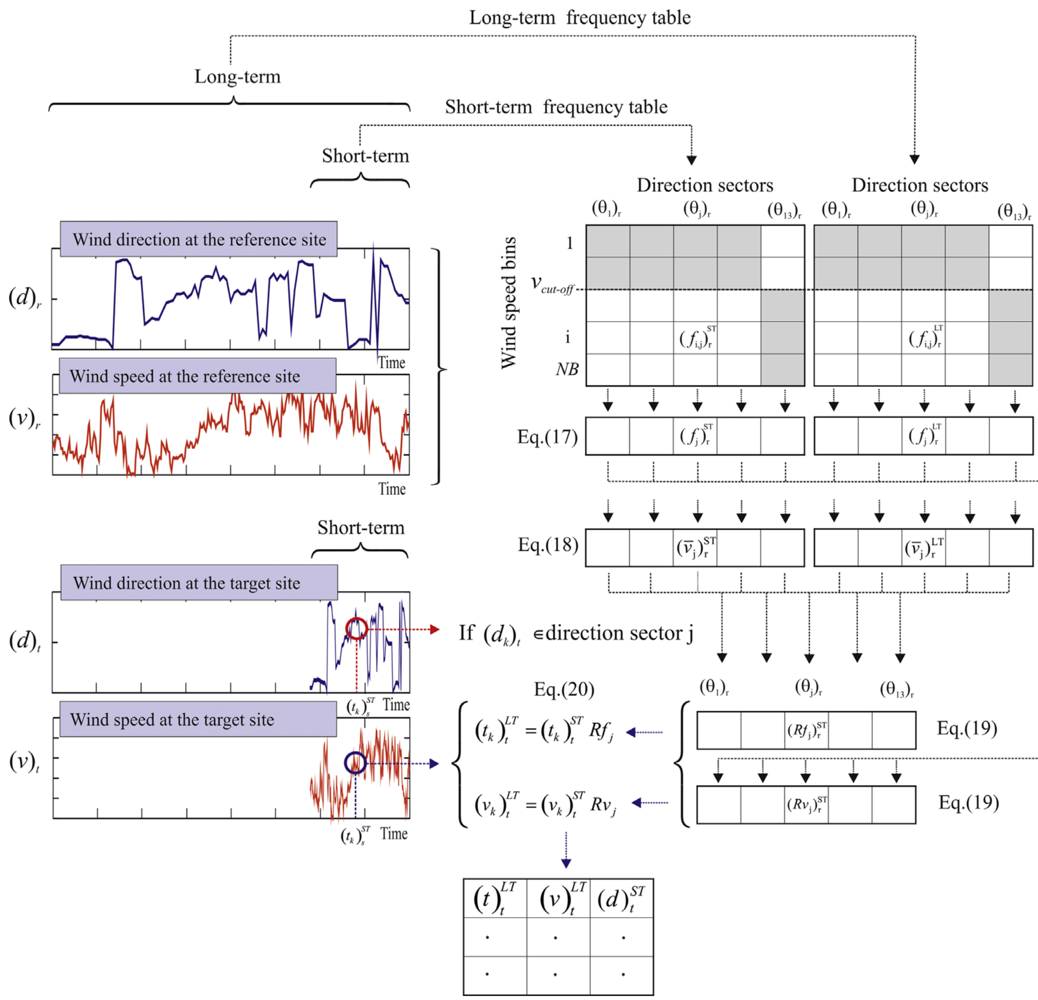

Fig. 9. Flow diagram created by the authors to illustrate the procedure used by the MSD method.

Using the generated wind speed and time data, together with the wind direction data recorded at the target site, a table of three columns is constructed, Fig. 9, representing the long-term wind data at the target site. Another table can be built from this of long-term wind speed and direction frequencies at the target site.

Wind speed and direction histograms can be configured from the table of long-term frequencies obtained with the MSD method to which continuous probability distributions can be fitted [144,145].

Hanslian [129] presented two ratios methods named the distribution of ratios method and the fingerprinting method. Both methods bin the wind speeds by the reference site wind direction. As a result, the first method provides an artificial series of wind data. The wind speed and direction are estimated independently. The second method does not provide a time series of data, but rather frequency distributions.

The methods of ratios described here assume that the wind direction at the target site is equivalent to that at the reference site.

4.4.2. Methods based on first-order linear regressions

By far the most commonly used MCP methods in the wind industry have been based on linear regression to characterise the

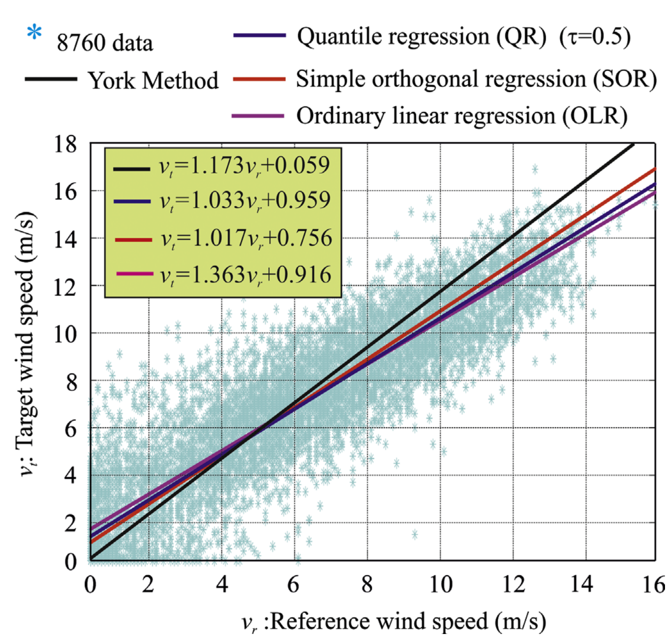

Fig. 10. Results of applying the OR, SOR, QR and York methods to the hourly mean wind data recorded at 10 m agl during 2006 at two anemometer stations installed on the island of Gran Canaria (Spain). The reference and target stations are named WS-6 and WS-5, respectively, in Ref. [102].

(*footnote continued*)
before the adjustment. However, if any frequency ratio has been adjusted to 1, as described in the main text, the two total times will not be equal.

relationship between the short-term (ST) wind speeds of the reference (r) and target (t) sites [128], followed by use of this relationship for long-term (LT) estimation of the wind speed time series at the target site:

$$(v_j)_t^{LT} = \beta (v_j)_r^{LT} + \alpha + \varepsilon_j \tag{21}$$

In Eq. (21), $(v_j)_t^{LT}$ are the estimated wind speeds for the target site, $(v_j)_r^{LT}$ are the wind speeds measured at the reference site and β and α are the slope and offset of the straight line which it is considered relates them. It is usually assumed that ε_j is white noise.[13] The statistical distribution of the so-called residuals, ε_j, can be estimated by using the short-term data, Eq. (22). These residuals should be Gaussian[14] and can be added to Eq. (21), estimating them with a random number generation method, such as the Monte Carlo method [147],[15]:

$$\varepsilon_j = (v_j)_t^{ST} - [\beta (v_j)_r^{ST} + \alpha] \tag{22}$$

If, when using Eq. (21), the predicted wind speed is negative this speed is taken as zero.

As described in Section 4.2, with the idea of taking into consideration the effects of wind direction the data are usually binned or grouped in wind direction sectors and the parameters of Eq. (21) are determined for each of the different sectors.

In order to determine the parameters or linear coefficients (β,α) of fit of the regression line, Eq. (21), use has fundamentally been made in the methods that have been proposed of the ordinary linear regression (OLR) [13,25–26,28,61,65,74,76,99,103,105–106,122–124,126–128] and, to a lesser extent, the orthogonal regression (OR) [25–26,28,65,76,128] and quantile regression (QR) [28].

In the OLR, the parameters β and α, Eq. (21), which best fit a straight line to the data, are determined by least squares. Given that α and β depend on wind speed and direction, n in Eq. (23) is the number of wind speeds in the direction sector of the reference site under analysis.

The aim of the OLR is to reduce to a minimum the sum of the squares of the vertical distances between the values of the dependent data (observed wind speeds at the target site) and the corresponding values as predicted by the fitted line [66,67], Fig. 10. The OLR assumes that the independent variable (reference site wind speed) is error free (that is, the values of the dependent variable are known exactly) and that only the dependent variable (target site wind speed) has an error component [148]. The basic conditions or fundamental hypotheses for this model are that the errors are homoscedastic variables (the errors have the same variance in all the observations of the endogenous variable) that the errors have a normal distribution with zero mean and that the pairs of sample observations, $[(v_j)_t, (v_j)_r]$, are independent of each other, which will mean that the covariance of the errors is zero ($\forall i,j,i \neq j$):

$$\beta = \frac{\sum_{j=1}^{n} (v_j)_r^{ST} (v_j)_t^{ST} - (1/n) \sum_{j=1}^{n} (v_j)_r^{ST} \sum_{j=1}^{n} (v_j)_t^{ST}}{\sum_{j=1}^{n} (v_j^2)_r^{ST} - (1/n) \left[\sum_{j=1}^{n} (v_j)_r^{ST} \right]^2} = \frac{(s_{v_r v_t})^{ST}}{(s_{v_r}^2)^{ST}}$$

$$\alpha = \bar{v}_t^{ST} - \beta \bar{v}_r^{ST} \tag{23}$$

In Eq. (23), $(v_j)_r^{ST}$ and $(v_j)_t^{ST}$ are the observed short-term (ST) wind speeds at the reference and target site, respectively. n is the number of wind speed data, $s_{v_r v_t}$ is the covariance between $(v_j)_r^{ST}$ and $(v_j)_t^{ST}$, and $s_{v_r}^2$ is the variance of $(v_j)_r^{ST}$.

Given that the independent variable is also measured and is therefore subject to errors, the use of orthogonal regression has been proposed [28,65,76,128] as OR assumes that both the dependent and independent variables have associated errors, εv_t and εv_r.

The aim in OR is to reduce to a minimum the orthogonal distances (perpendiculars) of the data points to the fitted line [148–150]. The linear model that results is known as the Major Axis [151].

As shown by Fuller [149], in the general orthogonal regression method (GOR), if the ratio λ, Eq. (24), between the variances of the errors of both variables is known from independent information, the best estimators of the coefficients β and α are those given by Eq. (25):

$$\lambda = \frac{s_{\varepsilon v_t}^2}{s_{\varepsilon v_r}^2} \tag{24}$$

$$\beta = \frac{(s_{v_t}^2)^{ST} - \lambda (s_{v_r}^2)^{ST} + \sqrt{[(s_{v_t}^2)^{ST} - \lambda (s_{v_r}^2)^{ST}]^2 + 4\lambda (s_{v_t v_r}^2)^{ST}}}{2 (s_{v_t v_r})^{ST}}$$

$$\alpha = \bar{v}_t^{ST} - \beta \bar{v}_r^{ST} \tag{25}$$

In Eq. (25), $s_{v_t}^2$ is the variance of $(v_j)_t^{ST}$.

In the simple orthogonal regression method (SOR), it is assumed that both variables have the same uncertainty ($\lambda = 1$). When λ cannot be reliably evaluated, Castellaro and Bormann [152] show that, under reasonable assumptions, the SOR (equivalent to the GOR with $\lambda = 1$) performs better than the OLR, Fig. 10.

Carroll and Rupper [153] argue that the slope, β, is often overestimated when ignoring the error of the equation, $\sigma_{\varepsilon Eq}$. So, they say that use of OR should include a careful assessment of the error of the equation, and not just the customary (often informal) estimation of the ratio of variances of the measurement errors of Eq. (27). That is, they propose the use of Eq. (26) instead of Eq. (24):

$$\lambda = \frac{s_{\varepsilon v_t}^2 + \sigma_{\varepsilon Eq}^2}{s_{\varepsilon v_r}^2} \tag{26}$$

When only the λ relationship is known, the GOR represents the simplest and most widely used procedure. However, there are other OR methods, such as the chi-square regression (CSQ) and the weighted total least squares (WLS) [148,154–155][16] which are preferable when the uncertainty of each observation is known [148]. As its name indicates, the CSQ is based on minimisation of the chi-square statistic, χ^2, defined as the sum of the squares of n random residuals, normalized by their variances:

$$\chi^2(\alpha, \beta) = \sum_{i=1}^{n} \frac{[(v_i)_t^{ST} - \alpha - \beta (v_i)_r^{ST}]}{(s_i^2)_r^{ST} + \beta^2 (s_i^2)_t^{ST}} \tag{27}$$

where $(s_i)_r^{ST}$ and $(s_i)_t^{ST}$ are, respectively, the standard deviations at the ith point of the short-term wind speeds of the reference and target sites, respectively.

In 1966, York [150] published an article showing how to perform the fit of the straight line to the data. York [150] showed that the equations could be resolved analytically. However, the

[13] According to Thøgersen et al. [24], in many cases the data show that a better assumption consists of modeling the residuals conditioned on both wind speed and direction. This model is available in WindPRO, where the residuals are modeled as Gaussian, but with the mean and standard deviation conditioned on the wind speed and wind direction.

[14] The distribution of the residuals should be examined for the purpose of checking the model. The type of examination almost always comprises easy to construct graphs which are generally very revealing when the assumptions are violated [66].

[15] Consideration of the residuals is important because wind energy varies with the cube of the wind speed and, therefore, the positive residuals contribute much more to increasing the energy than the negative residuals contribute to lowering it.

[16] The methods presented in [154,155] are based on minimization of the chi-square statistic, but the formulation differs from that indicated in Eq. (27). Krystek and Anton [155] proposed a model for when the uncertainties of the two variables are correlated.

analytical solution of the equations is based on an assumption with respect to the slope of the straight line, generally taking as the slope that obtained with the Major Axis method. York [150] suggests the use of an iterative routine to obtain the best estimation of the slope β, Fig. 10. York's method has been implemented in various software applications used in the wind industry [26,65,76] that have MCP modules. Ref. [75] contains a routine in Fortran 77 which facilitates calculation of the regression line using York's method. According to Rebbeck [55], RES (Renewable Energy Systems Ltd.) claimed that its method was equivalent to York's. However, this is not true, as RES uses an empirical technique which does not match the mathematical rigour of York's method.

As stated above, the aim in classical regression methods is to minimise the sum of the squares of the residuals and use the mean as estimator. However, quantile regression, introduced by Koenker and Bassett [156], is not based on assumptions as restrictive as those of classical linear regression and aims to model the relationship that exists between the dependent and independent variable for different quantiles (median or any other quantile) of the distribution of the dependent variable. As a result, QR is able to provide a much more complete statistical analysis of the relationships between random stochastic variables [157]. Estimation of the parameters in the case of quantile regression is carried out by minimisation of the asymmetrically weighted absolute deviations, Eq. (28), where τ is the quantile ($0\le\tau\le1$). As a result of the structure of QR, estimation of the parameters (α_τ,β_τ) cannot be reduced to a numerical linear algebra problem, as in the case of the minimum squares method, Fig. 10. In this case, a linear programming problem is involved which can be solved by the simplex method [75].

$$(\alpha_\tau,\beta_\tau)=\underset{\alpha_\tau,\beta_\tau\in\mathbb{R}^n}{\operatorname{argmin}}\left\{\sum_{(v_i)_t^{ST}\ge\alpha_\tau-\beta_\tau(v_i)_r^{ST}}\tau|(v_i)_t^{ST}-\alpha_\tau-\beta_\tau(v_i)_r^{ST}|\right.$$
$$\left.+\sum_{(v_i)_t^{ST}<\alpha_\tau-\beta_\tau(v_i)_r^{ST}}(1-\tau)|(v_i)_t^{ST}-\alpha_\tau-\beta_\tau(v_i)_r^{ST}|\right\} \tag{28}$$

It should be mentioned that among the above-described ordinary regression, orthogonal regression and quantile regression models with two parameters (α,β,), there are models that have been proposed which are conditioned on the straight line passing through the offset [2,25,28,65,98,105]. That is, the offset (α) of the regression line has a zero value. So, a perfect correlation is considered to exist between the data of the target and reference stations. According to Joensen et al. [132], from a physical point of view it is not logical to include the offset, as this implies that the wind speed of the target site will not be zero when the wind speed at the reference site is zero. However, in terms of the mean square error, performance increases when the offset is included, as the model is only approximate. According to Jain [2], though regression through (0,0) is intuitive, it does not provide good representation for the higher wind speeds, which are precisely the winds which result in energy production. King and Hurley [98] consider that a straight line through the origin does not provide a good model for relating wind speed distributions. Thøgersen et al. [24] report that the MCP module of the WinPro software application can force the regression line to pass through the origin (0,0). However, they state that this option should only be used with care, as in general it gives a significantly poorer fit of the data than methods in which a non-zero intersection with the Y-axis is permitted.

Beltrán et al. [99] propose replacing the linear regression method with a linear algorithm which they called the bin or bin-fitting method [158], as it is based on the standard procedure used in measurement of the power curve of a wind turbine [159].

In the proposed algorithm [99], the wind speed data are grouped into bins of wind speed and wind direction sectors. In this case, the wind speeds of the target station are binned versus the binned measured wind speeds of the reference station in ranges of 0.5 m/s. In each bin b, where the number of data must be higher than 10, the means of the wind speeds of the reference and target stations are calculated. A set of points is thus obtained in a Cartesian system in which the x-axis represents the wind speed at the reference station and the Y-axis the wind speed at the target station. Then, by means of linear interpolation between these points, the following equation, the wind speeds at the target station are estimated:

$$(v_i)_t^{LT}=(\overline{v}_b)_t^{ST}+\left[(v_i)_r^{LT}-(\overline{v}_b)_r^{ST}\right]\frac{(\overline{v}_{b+1})_t^{ST}-(\overline{v}_b)_t^{ST}}{(\overline{v}_{b+1})_r^{ST}-(\overline{v}_b)_r^{ST}} \tag{29}$$

where $(v_i)_t^{LT}$ is the long-term wind speed at the target station, $(\overline{v}_b)_t^{ST}$ is the short-term mean wind speed of bin b at the target station, $(\overline{v}_b)_r^{ST}$ is the short-term mean wind speed of bin b at the reference station and $(v_i)_r^{LT}$ is the long-term wind speed pertaining to the range included between the mean speeds of bins b and b+1 at the reference station.

In those cases in which the long-term reference station wind speeds are greater than the highest bin of the interpolation, estimation of the long-term target station wind speed is carried out through a linear function. The authors propose that this linear function be obtained through a fit to the set of interpolation points $[(\overline{v}_b)_r^{ST}, (\overline{v}_b)_t^{ST}]$, excluding data below 3.25 m/s, Fig. 11.

LeBlanc et al. [128] proposed a method similar to the bin method, which they call the Vertical Slice method. The wind speed domain is split into vertical bands or slices and a piecewise linear fit is performed of the mean of the dependent variable of each slice or bin. If the number of data in a particular bin or slice is too low to obtain the mean of the dependent variable, then the fit is performed using a standard linear regression. This method has been implemented in the WindoGrapher software [25].

King and Hurley [98] have proposed three MCP methods, named SpeedSort, DynaSort and Scatter. These have been incorporated in the set of MCP methods used by WindoGrapher [25].

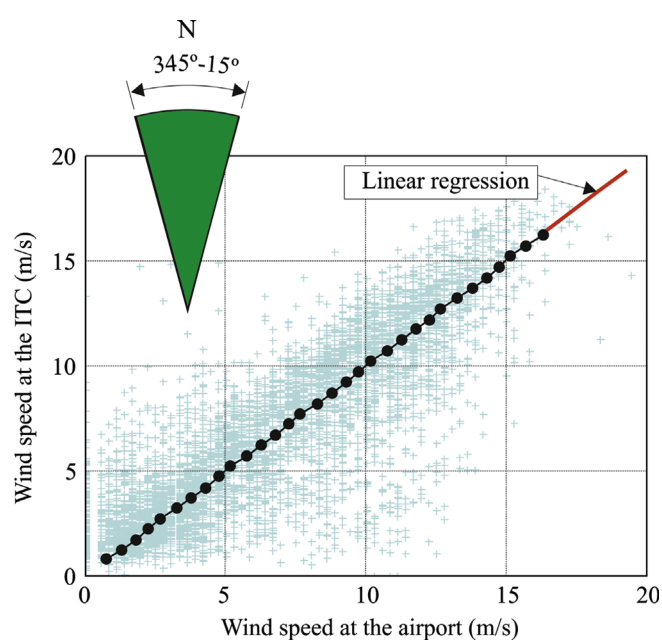

Fig. 11. Results of applying the method of bins to the hourly mean data recorded at 10 m agl during 2010 at two anemometer stations (Airport and ITC) installed on the island of Gran Canaria (Spain).

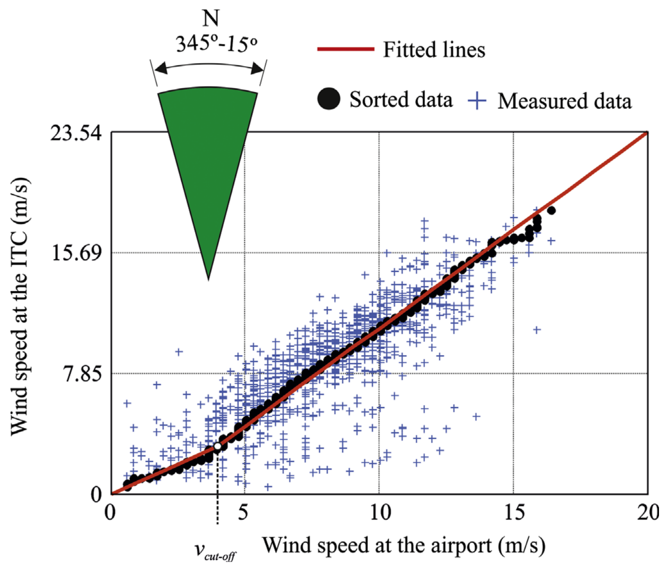

Fig. 12. Results of applying the SpeedSort method to the hourly mean data recorded at 10 m agl during 2010 at two anemometer stations (Airport and ITC) installed on the island of Gran Canaria (Spain).

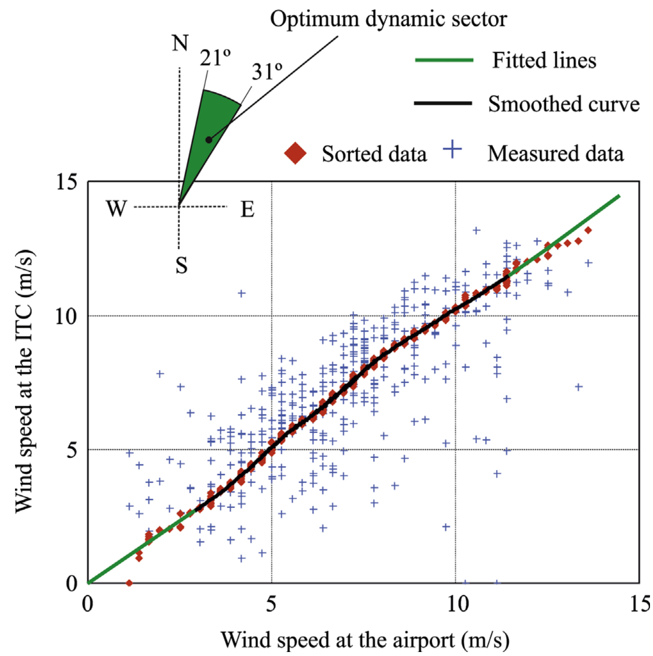

Fig. 13. Results of applying the DynaSort method to the hourly mean data recorded at 10 m agl during 2010 at two anemometer stations (Airport and ITC) installed on the island of Gran Canaria (Spain).

The algorithm of the SpeedSort method, in an initial stage, assigns each wind speed of the time series representing the short-term period to one of the standard twelve 30° direction sectors defined by the reference site wind direction. The wind speeds of the reference site below the calm threshold (1 m/s) are randomly distributed among the direction sectors. In each sector, the target and reference wind speeds are independently ordered (in ascending or descending sequence), Fig. 12. In this respect, the SpeedSort method displays a certain similarity to the method of bins [99]. In a second stage, two linear relationships are proposed in principle to characterise in each direction sector the relationship between the ordered short-term wind speeds of the target and reference sites. Each of these linear relationships represents a range of wind speeds of the reference site. These lines of fit try to estimate the relationship that exists between the wind speed frequency distributions of the two sites as opposed to the relationship between values of the time series of their wind speeds.

The authors [98] propose using a reference site cut-off speed, $v_{cut-off}$, to establish the separation of the two datasets which will be represented by these linear functions, which they take as the lower value of $\overline{v}_r^{LT}/2$ and 4 m/s.[17] They use a linear regression of parameters α and β, Eq. (24),[18] to characterise wind speeds higher than the cut-off speed in a given sector if the number of data in that sector at least reaches a critical value which depends on the correlation coefficient between the target and reference site wind speeds for the same sector. If this condition is not met, but there is at least an established minimum number of data, then the authors propose using a straight line which passes through the origin and the centroid of the data. If this final restriction is not met either, they propose using a straight line which passes through the origin and which has a slope of 45°. That is, in this last case, the wind speeds at the target site are considered to be the same as those at the reference site.

To characterise the relationship between wind speeds lower than the cut-off point, a straight line is proposed which starts from the lower end of the straight line obtained with Eq. (21) and which passes through the origin. That is, a straight line of zero offset and slope β_2:

$$(v_j)_t^{LT} = \beta_2 (v_j)_r^{LT} + \varepsilon_j = \left[\beta + \frac{\alpha}{v_{cut-off}}\right](v_j)_r^{LT} + \varepsilon_j \tag{30}$$

To estimate the parameters α and β of the regression line, the authors [98] propose using a variation of the OLR. In the description of the procedure, instead of minimising the sum of the squares of the distances perpendicular to the x-axis, they minimise the sum of the squares of the distances perpendicular to the line that passes through the origin and centroid of wind speeds higher than the cut-off speed of the reference site.[19]

According to the authors, SpeedSort is better than the EWEA method in terms of representation of wind speed frequencies. However, they admit that a straight line generally does not perfectly represent the distributions and therefore gives rise to scatter. In this method, an analysis is also undertaken by the authors [98] of the differences between the wind direction of the target and reference sites for each data of the time series, and the long-term wind direction at the target site is predicted as explained in Section 4.6 of this review. The output of results generated by the algorithm of the SpeedSort method comprises a table of relative frequencies, representing the long-term wind speeds and directions at the target site.

The DynaSort method differs from SpeedSort in three main ways. First, DynaSort uses a number N of direction sectors which is not fixed[20] but is dependent on the number of data filtered in a previous process and whose bounds are chosen by ensuring that all the sectors have an equal number of data.[21] The algorithm of

[17] According to the authors [98], these values were chosen with the idea of optimizing the fit of the line to the range of wind speeds which is of interest from the point of view of a wind turbine's electricity production.

[18] If the linear relationships estimate a negative long-term wind speed at the target site, this wind speed is assigned a value of zero.

[19] According to the authors [98], there are two advantages to this procedure: (a) it gives the same line independently of which site is on the vertical axis and (b) any scattering will tend to pull the regression line towards the line that passes through the origin and centroid rather than towards a horizontal line that passes through the centroid.

[20] For convenience sake, they limit this number to 24.

[21] The authors [98] have programmed the algorithm of this method to solve cases in which there are surplus data.

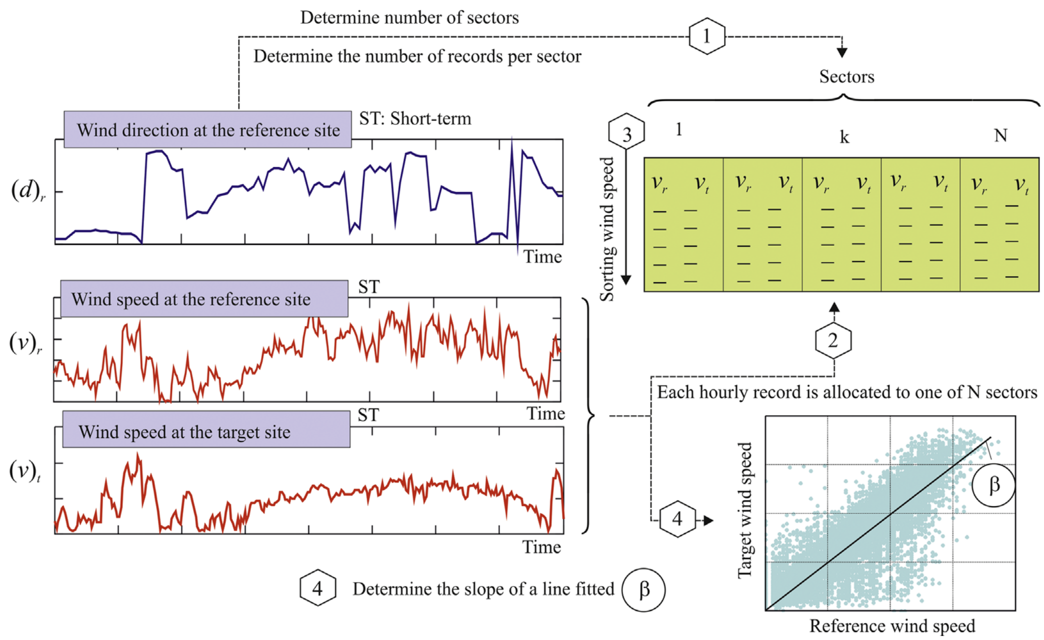

Fig. 14. Flow diagram created by the authors to facilitate interpretation of the first block of tasks of the algorithm followed by the scatter method.

the DynaSort method proceeds in a similar way to that of SpeedSort, except that instead of assigning each wind speed of the time series representing the short-term period to one of 12 fixed direction sectors, it assigns it to one of N dynamically calculated direction sectors defined by the reference site wind direction. Second, the DynaSort method, in addition to independently sorting in each sector of each site all the wind speeds (that is, without carrying out a prior filtering), also sorts the wind directions (though in this case using filtered data) at the reference and target sites for subsequent prediction of the long-term target site wind direction, as explained in Section 4.6. As occurs with the wind speeds, one consequence that results from sorting the wind directions is that the relationships that are established are based on wind direction frequency distributions instead of being based on time series of the wind directions.

The third difference is seen in that the algorithm of the DynaSort method does not fit straight lines to the n sorted wind speeds in each sector. In this method, the algorithm tries to reduce the amount of variation present in the set of ordered wind speed data by smoothing it out. For this purpose, the concept of moving averages is employed [67] with a span of M data, Eq. (31), and the trend line is then used to represent the relationship between the long-term target and reference site wind speeds, Fig. 13:

$$(v'_{k+1})_t^{ST} = \frac{\sum_{j=1+k}^{M+k}(v_j)_t^{ST}}{M};\ (v'_{k+1})_r^{ST} = \frac{\sum_{j=1+k}^{M+k}(v_j)_r^{ST}}{M};\ k=0\ldots n-M \quad (31)$$

If the long-term reference site wind speed is within the range covered by the smoothed curve, the long-term target site wind speed is estimated from this curve, interpolating if necessary. One of the drawbacks of this technique of moving averages is that the data of the start and end of the sorted series are lost. So, the authors [98] propose taking a straight line that goes from the first moving average to the offset to estimate the long-term target site wind speed if the long-term reference site wind speed is lower than the range covered by the smoothed curve. However, if the long-term reference site wind speed is higher than the range covered by the smoothed curve, the authors [98] propose using a straight line that commences at the last moving average and has a slope β_3, Eq. (32), to estimate the long-term target site wind speed. This slope is defined by the bisector of the angle formed between a straight line that passes through the origin and the last moving average and a straight line of 45° slope:

$$\beta_3 = \tan\left\{\frac{45}{2} + \frac{1}{2} a\ \tan\left[\frac{(v'_{n-M+1})_t^{ST}}{(v'_{n-M+1})_r^{ST}}\right]\right\} \quad (32)$$

According to the authors [98], the results of the studies they carried out show that the DynaSort method was slightly less accurate than SpeedSort for periods below one year, though DynaSort did tend to improve its performance compared to SpeedSort as the length of the short-term period grew. When the short-term period is not very long the irregularities in the adjusted curve are perhaps more due to random wind variations between the sites than to any long-term effect. This suggests to the authors [98] that precise representation of wind speed frequency distributions requires a short-term period that covers at least one complete year of data.

Conceptually speaking, the Scatter method differs from SpeedSort and DynaSort in that in its procedure the long-term reference site data are modified to obtain the long-term target site data by using single data of the time series of the short-term period of both sites. The algorithm of the Scatter method contains two blocks of differentiated tasks. In the first block, the data of the period corresponding to the short-term period are distributed, based on the reference site wind direction, among the N dynamically chosen direction sectors[22] which are determined according to the total number of data of the period and whose bounds are defined by ensuring that these sectors have an equal number of data,[23] Fig. 14. Then, the wind speeds are independently sorted in ascending or descending order in all the sectors of both sites.

In the second block, Fig. 15, the data of the time series of the long-term wind speeds and directions are processed. For each data of the time series of the long-term reference site wind speeds,

[22] The wind speeds of the reference site below the calm threshold are randomly distributed among all the direction sectors.

[23] The authors [98] programmed an algorithm different to that introduced in the DynaSort method to solve cases in which there are surplus data.

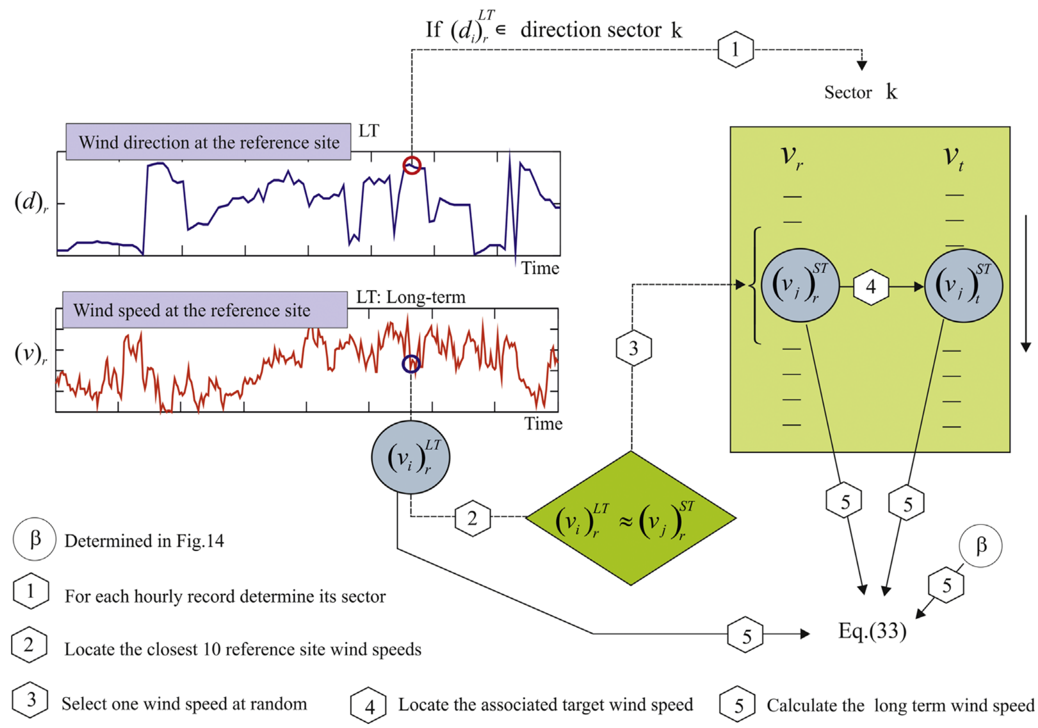

Fig. 15. Flow diagram created by the authors to facilitate interpretation of the second block of tasks of the algorithm followed by the scatter method.

$(v_j)_r^{LT}$, the algorithm that corresponds to this block performs a set of routines. The direction sector, $(\theta_k)_r$, is determined from among the sectors defined in the first block to which it must belong as a function of the wind direction, $(d_j)_r^{LT}$, associated to the wind speed, $(v_j)_r^{LT}$, in the time series. The ten closest wind speeds to $(v_j)_r^{LT}$ are located from the set of reference station wind speeds of this sector, $(\theta_k)_r$, in the short-term period and one of them,$(v_i)_r^{ST}$, is randomly selected. Following this, the corresponding wind speed,$(v_i)_t^{ST}$, in sector $(\theta_k)_t$ of the target site is then taken. The next line of the routine estimates the long-term target site wind speed through the following equation[24]:

$$(v_j)_t^{LT} = (v_i)_t^{ST} + [(v_j)_r^{LT} - (v_i)_r^{ST}]\beta \tag{33}$$

In Eq. (33), β is the slope of a straight line, the parameters of which are determined through a linear regression, Eq. (21), which in this case characterises the relationship between the wind speeds of all the target site sectors in function of the short-term wind speeds of all the reference site sectors. The output of results generated by the algorithm of the Scatter method consists of a table of relative frequencies, representing the long-term target site wind speeds and directions. For the construction of this table the algorithm uses the wind speed bins versus the standard 12 wind direction sectors.

Rogers et al. [106] propose using the so-called variance ratio method, Eq. (34), to predict the long-term wind speed at a target site:

$$(v_j)_t^{LT} = \left[\bar{v}_t^{ST} - \left(\frac{s_t^{ST}}{s_r^{ST}}\right)\bar{v}_r^{ST}\right] + \left(\frac{s_t^{ST}}{s_r^{ST}}\right)(v_j)_r^{LT} \tag{34}$$

The slope and offset of the linear relationship given by Eq. (34) are determined by Rogers et al. [106] forcing the mean and variance of the estimated long-term wind speeds at the target site, during the short concurrent data period, to coincide with the mean and variance of the observed wind speeds at the same site.

[24] If the estimated wind speed is negative, it is set to zero.

This mean that the square of the slope β of the regression line coincides with the ratio between the variances of the target site and reference site wind speeds in the concurrent data period. Rogers et al. [106] showed that the variance ratio method gives better results than ordinary linear regression when predicting mean wind speed. As in the methods mentioned above based on first-order linear regressions with non-zero offset, if the wind speeds predicted with Eq. (34) are negative then they are considered to have zero value. This method has been implemented in the MCP module of the WindoGrapher software [25].

Achberger [160] proposes decomposing the reference and target site wind speeds with two orthogonal axes (*X*,*Y*):

$$\begin{aligned} v_x &= v\ \sin(d) \\ v_y &= v\ \cos(d) \end{aligned} \tag{35}$$

where v is the magnitude of the observed wind speed, d is the wind direction and v_x and v_y are the Cartesian wind components. West-east and south-north, respectively, are usually taken as the positive directions of the *X*- and *Y*-axes [145].

Two independent linear regressions are then performed; one regression between the components on the *X*-axis of the reference and target site short-term wind speeds, and another regression between the same components on the *Y*-axis. Based on the regressions performed between the wind speed components, the slopes (β_x,β_y) and offsets (α_x,α_y) are determined and the components of the long-term target site wind speeds are estimated:

$$[(v_j)_x]_t^{LT} = \beta_x[(v_j)_x]_r^{LT} + \alpha_x + \varepsilon_{x,j} \tag{36}$$

$$[(v_j)_y]_t^{LT} = \beta_y[(v_j)_y]_r^{LT} + \alpha_y + \varepsilon_{y,j} \tag{37}$$

In a final step, the long-term target site wind speeds are estimated by combining their components:

$$(v_j)_t^{LT} = \{[[(v_j)_x]_t^{LT}]^2 + [[(v_j)_y]_t^{LT}]^2\}^{1/2} + \varepsilon_j \tag{38}$$

It can be deduced from Eq. (38) that the wind speeds predicted with this method cannot take negative values.

Nielsen et al. [74] propose use of a vector linear relationship. Using this model, the components (in a system of X, Y orthogonal coordinates) of the long-term target site wind speeds are obtained through a multilinear regression, Eq. (39). This method is referred to as the “vector” model [74] or “vector” method [106]:

$$\begin{bmatrix}[(v_j)_x]_t^{LT}\\ [(v_j)_y]_t^{LT}\end{bmatrix}=\begin{bmatrix}\alpha_1\\ \alpha_2\end{bmatrix}+\begin{bmatrix}\beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{bmatrix}\begin{bmatrix}[(v_j)_x]_r^{LT}\\ [(v_j)_y]_r^{LT}\end{bmatrix} \tag{39}$$

The coefficients $\beta_{\rm ij}$ and $\alpha_{\rm j}$ are determined through the matrix product shown in the following equation:

$$\begin{bmatrix}\alpha_1 & \alpha_2\\ \beta_{11} & \beta_{12}\\ \beta_{21} & \beta_{22}\end{bmatrix}=\begin{bmatrix} n & \sum_{j=1}^{n}[(v_j)_x]_r^{ST} & \sum_{i=1}^{n}[(v_i)_y]_r^{ST}\\ \sum_{j=1}^{n}[(v_j)_x]_r^{ST} & \sum_{j=1}^{n}\{(v_j)_x^2\}_r^{ST} & \sum_{i=1}^{n}\{(v_i)_y\}_r^{ST}\{(v_i)_x\}_r^{ST}\\ \sum_{j=1}^{n}\{(v_j)_y\}_r^{ST} & \sum_{j=1}^{n}\{(v_j)_y\}_r^{ST}\{(v_j)_x\}_r^{ST} & \sum_{i=1}^{n}\{(v_i)_y^2\}_r^{ST}\end{bmatrix}^{-1}$$
$$\times\begin{bmatrix}\sum_{j=1}^{n}\{(v_j)_x\}_t^{ST} & \sum_{j=1}^{n}\{(v_j)_y\}_t^{ST}\\ \sum_{j=1}^{n}\{(v_j)_x\}_t^{ST}\{(v_t)_x\}_r^{ST} & \sum_{j=1}^{n}\{(v_j)_y\}_t^{ST}\{(v_j)_x\}_r^{ST}\\ \sum_{j=1}^{n}\{(v_j)_x\}_t^{ST}\{(v_j)_y\}_r^{ST} & \sum_{j=1}^{n}\{(v_j)_y\}_t^{ST}\{(v_j)_y\}_r^{ST}\end{bmatrix} \tag{40}$$

The modules of each of the estimated long-term target site wind speeds are determined through Eq. (38).

Achberger [160] also proposes a model based on a vector regression technique [161]. First, the model considers the two horizontal components of the wind speed as the real and imaginary parts of a complex number:

$$\begin{aligned}v_t&=(v_x)_t+i(v_y)_t\\ v_r&=(v_x)_r+i(v_y)_r\end{aligned} \tag{41}$$

where $i=\sqrt{-1}$ and the variances and covariance are defined by the following equations, respectively:

$$(s_{v_r})^2=[s_{(v_x)r}]^2+[s_{(v_y)r}]^2 \tag{42}$$

$$(s_{v_t})^2=[s_{(v_x)t}]^2+[s_{(v_y)t}]^2 \tag{43}$$

$$s_{v_rv_t}=[s_{(v_x)_r(v_x)t}+s_{(v_y)_r(v_y)t}]+i[s_{(v_x)_r(v_y)t}-s_{(v_y)_r(v_x)t}] \tag{44}$$

Using this model, the long-term target site wind speeds are obtained through the following equation:

$$\begin{aligned}(v_j)_t^{LT}&=[(v_j)_x]_t^{LT}+i[(v_j)_y]_t^{LT}\\ &=(\alpha_x+i\alpha_y)+(\beta_x+i\beta_y)\{[(v_j)_x]_r^{LT}+i[(v_j)_y]_r^{LT}\}\end{aligned} \tag{45}$$

The parameters β_x and β_y are determined through Eqs. (46) and (47), respectively:

$$\beta_x=(\rho_{v_rv_t})_x\frac{s_{v_t}}{s_{v_r}} \tag{46}$$

$$\beta_y=(\rho_{v_rv_t})_y\frac{s_{v_t}}{s_{v_r}} \tag{47}$$

In Eqs. (46) and (47), $(\rho_{v_rv_t})_x$ and $(\rho_{v_rv_t})_y$ constitute the real and imaginary components of the correlation vector $\rho_{v_rv_t}$. These components of the correlation vector are given by Eqs. (48) and (49), respectively:

$$(\rho_{v_rv_t})_x=\frac{s_{(v_x)_r(v_x)t}+s_{(v_y)_r(v_y)t}}{s_{v_r}s_{v_t}} \tag{48}$$

$$(\rho_{v_rv_s})_x=\frac{s_{(v_x)_r(v_y)s}-s_{(v_y)_r(v_x)s}}{s_{v_r}s_{v_s}} \tag{49}$$

The parameters α_x and α_y are determined through Eq. (50) and Eq. (51), respectively:

$$\alpha_x=[\overline{v}_x]_t^{ST}-\beta_x[\overline{v}_x]_r^{ST}+\beta_y[\overline{v}_y]_r^{ST} \tag{50}$$

$$\alpha_y=[\overline{v}_y]_t^{ST}-\beta_x[\overline{v}_y]_r^{ST}-\beta_y[\overline{v}_x]_r^{ST} \tag{51}$$

The modules of each of the estimated long-term target site wind speeds are determined through Eq. (38).

All the linear-regression-based methods mentioned above assume, with the odd exception [74,160] that the target site wind direction is equivalent to that of the reference site. However, Woods and Watson [101] point out that there are places where, for example, the complexity of the terrain often makes invalid the assumption that there exists a one-to-one correspondence between the reference and target site wind direction sectors. In this context, the authors propose a general approach which they call the matrix method that allows to take into consideration the possibility that the reference site wind rose is different to that of the target site. First, in the methodology they propose, a matrix **E** ($N\times N$) is created with the wind data binned in N[25] direction sectors of the reference and target sites during the concurrent data period. An element $e_{i,j}$ of the matrix **E** contains the number of times that the wind has blown simultaneously in sector i of the target site and in sector j of the reference site, Table 1. Subsequently, those elements, $e_{i,j}$, are discarded which comprise a small fraction, δ,[26] of the total number of data recorded in each sector. So, the work restarts with a new matrix **E**′ whose elements $e'_{i,j}$ are null or coincide with those of matrix **E** when the restriction given by the following equation is met, Table 2:

$$\frac{e_{i,j}}{\sum_{k=1}^{N}e_{i,k}}100>\delta\rightarrow e'_{i,j}=e_{i,j};\ i=1\ldots N;\ j=1\ldots N \tag{52}$$

From **E**′, a new matrix **Z** ($N\times N$) is constructed whose elements $z_{i,j}$ are expressed as a percentage of the total number of data of each sector i of the target site, Eq. (53), such that Eq. (54) is met for each row (sector) i, Table 3.

$$z_{i,j}=\frac{e'_{i,j}}{\sum_{k=1}^{12}e'_{k,j}}100;\ i=1\ldots N;\ j=1\ldots N \tag{53}$$

$$\sum_{j=1}^{N}z_{i,j}=1 \tag{54}$$

Once the matrix **Z** has been constructed, there are two approaches to predict the long-term wind speeds in each sector i of the target site.

In one of the approaches, the mean wind speed in each bin of a particular direction sector is considered to be equivalent to the overall mean of the wind speed of that sector. In addition, it is assumed that the regression line which is obtained with the wind speeds of a sector can be used in each of the individual bins of the sector. In this approach, the mean long-term wind speeds, $(\overline{v}_i)_t^{LT}$, are estimated for each sector i of the target site from the mean long-term wind speeds, $(\overline{v}_j)_r^{LT}$, of sectors j of the reference station by means of the weighting of first-order linear regressions obtained for the direction sectors j, Eq. (55). The weighting coefficients are composed of the elements $z_{i,j}$ of the matrix **Z**. β_j and α_j in Eq. (55) are the slope and offset, respectively, of the straight line that characterises the relationship between all the $e'_{i,j}$ ($i=1\ldots N$) short-term wind speed data of the reference site (r)

[25] In their work, the authors use 12 direction sectors ($N=12$).

[26] The authors state that a cut-off level δ of around 5–7% seems to give the best results. However, they also state that the choice of this cut-off percentage results from an examination of typical sites in the United Kingdom, but that there are possibly cases for which different cut-off levels should be adopted.

Table 1
Matrix **E** created with hourly mean data recorded during 2010 at two anemometer stations installed in the Canary Archipelago (Spain)[a] (Fig. 2).

Target site sectors (ITC)	Reference site sectors (Gran Canaria airport)											
	1	2	3	4	5	6	7	8	9	10	11	12
1	297	3								19	119	474
2	3576	233	12	6	7	4	9	12	22	25	85	888
3	125	251	67	22	17	4	6	4	9	5	21	46
4	66	17	31	31	26	8	3	4	2	7	13	34
5	54	10	10	13	36	30	7	8	3	7	14	28
6	63	12	6	10	18	24	9	10	2	9	20	52
7	90	8	6	5	17	44	64	18	6	8	19	40
8	131	37	33	14	17	56	257	266	99	51	54	36
9	30	5	2	2		4	5	24	43	28	30	15
10	8							1	15	17	12	5
11	2							2	6	21	11	3
12									2	11	3	1

[a] The 12 wind direction sectors were all taken of the same size (30°) and were measured in a clockwise direction. Sector 1 is defined by the bounds 0(N)–30° and sector 12 by the bounds 330–360°(N).

Table 2
Matrix **E**′ created from the frequencies seen in matrix **E** (Table 1), after rejecting the elements that comprise a fraction less than or equal to 5% of the total number of data recorded in each sector of the target site.

Target site sectors (ITC)	Reference site sectors (Gran Canaria airport)											
	1	2	3	4	5	6	7	8	9	10	11	12
1	297										119	474
2	3576											888
3	125	251	67									46
4	66	17	31	31	26						13	34
5	54			13	36	30					14	28
6	63	12			18	24					20	52
7	90				17	44	64	18			19	40
8	131					56	257	266	99		54	
9	30							24	43	28	30	15
10	8								15	17	12	5
11									6	21	11	3
12									2	11	3	1

Table 3
Matrix **Z** created from matrix **E**′ (Table 2) and in which each element is expressed as a percentage of the total number of data of each target site sector.

Target site sectors (ITC)	Reference site sectors (Gran Canaria airport)											
	1	2	3	4	5	6	7	8	9	10	11	12
1	33.4										13.4	53.3
2	80.1											19.9
3	25.6	51.3	13.7									9.4
4	30.3	7.8	14.2	14.2	11.9						6.0	15.6
5	30.9			7.4	20.6	17.1					8.0	16.0
6	33.3	6.3			9.5	12.7					10.6	27.5
7	30.8				5.8	15.1	21.9	6.2			6.5	13.7
8	15.2					6.5	29.8	30.8	11.5		6.3	
9	17.6							14.1	25.3	16.5	17.6	8.8
10	14.0								26.3	29.8	21.1	8.8
11									14.6	51.2	26.8	7.3
12									11.8	64.7	17.6	5.9

and target site (t) which pertain to the direction sector j:

$$(\overline{v}_i)_t^{LT} = \frac{\sum_{j=1}^{N} z_{i,j}[\beta_j(\overline{v}_j)_r^{LT} + \alpha_j]}{100};\ i = 1\ldots N \tag{55}$$

In the other approach, it is only assumed that the mean wind speeds of the bins of a sector are equivalent to the mean wind speed of that sector. So, in this approach the mean long-term wind speeds, $(\overline{v}_i)_t^{LT}$, are estimated for each sector i of the target site from the mean long-term wind speeds, $(\overline{v}_j)_r^{LT}$, of sectors j of the reference station through a first-order linear regression obtained for the direction sector i, Eq. (56). In this case, the mean long-term wind speeds in the N reference site direction sectors are weighted. The weighting coefficients are composed of the elements $z_{i,j}$ of the matrix **Z**. β_i and α_i in Eq. (56) are the slope and offset, respectively, of the straight line that characterises the relationship between all the $d'_{i,j}$ ($j = 1\ldots N$) short-term wind speed data of the reference site (r) and the target site (t) which pertain to the direction sector i:

$$(\overline{v}_i)_t^{LT} = \beta_i \frac{\sum_{j=1}^{N} z_{i,j}(\overline{v}_j)_r^{LT}}{100} + \alpha_i;\ i = 1\ldots N \tag{56}$$

The authors [101] state that, in general, when the correlations were good they found little difference between the hindcasts made with the two approaches. However, when the correlations were poor, better results were obtained with the second approach. The authors also undertook a comparison of their method with a standard MCP method of linear regression and a standard MCP method that included correction for changes of direction in a pair of sites where the data for approximately 100 days of measurement that they used showed influence of the so-called "channeling effect". The results indicate that hindcasts of the global mean of the target site obtained by the three methods were almost identical to the actual measurement. However, the method proposed by the authors obtained better hindcasts by direction sector than the other two methods used for purposes of comparison.

Vermeulen et al. [162] tested the applicability of the method proposed by Woods and Watson [101] to the studies they carried out on wind energy sites in Armenia. This work was undertaken within the framework of an Armenian–Dutch project called ArmNedWind, the purpose of which was to study the large-scale integration of wind energy in Armenia´s electricity grid. Measurements were taken over a year-long period at five sites in Armenia. Monitoring masts 50 m agl were used for this purpose (with sensors placed at 35 m agl). In reference [162], only two pairs of sites were analysed: Pushkin Pass (PP)–Karakhach Pass (KH) and PP–Lake Sevan (SE). KH and PP are situated in valleys in the Bazum mountain range and have similar topographic characteristics. SE is found on the shores of Lake Sevan (Ardanish). In this study, a matrix of 16 direction sectors ($N = 16$) was used instead of the $N = 12$ direction sectors studied in [101]. In addition, in this study they used regression lines which pass through the origin or, in other words, regression lines with zero offset value. In view of the results that were obtained, the authors concluded that the matrix MCP method of Woods and Watson [101] was applicable to sites in complex mountainous terrain with strong wind direction changes. However, they state that though the reasoning behind the proposal of Woods and Watson [101] to eliminate all bins with a frequency of occurrence below a certain cut-off level was plausible, its real-life application introduces a number of problems with solutions that are only arbitrary. With respect to the cut-off level, Vermeulen et al. [162] state that though it is possible that elements (bins) of the matrix **E** with a very low occurrence have little meaning, the actual choice of 'very low' level is completely arbitrary and could easily lead to the elimination of infrequent but significant bins. In the case of the PP-KH sites that were analysed [162], using cut-off levels of 0.1%, 1% and 5% in the matrix **E** implies data losses of 6%, 31% and 74%, respectively. Also, elimination of data of the matrix **E** implies that the mean of the remaining data contained in the matrix **E**' must be adjusted. This adjustment or compensation is carried out by proportionally increasing the frequency of occurrence of the remaining data, Eq. (53). However, the authors point out that this assumption is again arbitrary, as the

eliminated wind speeds are not known. Nevertheless, not applying a filter will result in a certain number of bins with no meaning.

Vermuelen et al. [162], to avoid the use of bins with no meaning, propose checking the level of correlation between the bins and ignoring the entire matrix if very few bins show a significant correlation. This was the case of the PP–SE pair, for which the data were related in just one portion of the time (30%) and, consequently, it was not possible to convert the mean short-term wind speed of SE into a long-term estimation using the data of PP.

4.4.3. Higher than first-order linear methods

Though MCP methods based on first-order linear regression tools have been the most commonly used methods, other types of linear function have been proposed. In the literature [55,106–107,130,132,163] and in wind industry software [24,28,58,105], regressions have also been developed using higher order polynomials.

Deane et al. [163] propose a method that they call the fragmented interaction regression method to fill the gaps in the recorded wind speed data at a target site. Like the traditional MCP linear regression, this method splits the wind direction into sectors but, instead of using a simple linear regression between the target and reference sites it uses a multilinear regression which, in addition to considering the wind speed and direction, also gives a value to the product of both variables. So, for each direction sector, the estimated short-term wind speeds of the target site, $(v_j)_t^{ST}$, are given by the following equation:

$$(v_j)_t^{ST} = \beta_1 (v_j)_r^{ST} + \beta_2 (d_j)_r^{ST} + \beta_4 (v_j)_r^{ST} (d_j)_r^{ST} + \alpha + \varepsilon_j \tag{57}$$

In Eq. (57), $(d_j)_r^{ST}$ and $(v_j)_r^{ST}$ are, respectively, the observed reference site wind directions and speeds during the concurrent data period. β_1 and β_2 are the linear coefficients, β_4 is the coefficient of interaction and α is the offset.

The authors [163] conclude that the fragmented interaction regression is an improvement over the linear regression. However, they also state that more research is required to make proper use of the results of their study.

Joensen et al. [132] state that if the measurements of the target and reference site are not taken at the same height agl, then a first-order linear model is not a reasonable choice. In these cases, they propose using two alternative models to take into consideration atmospheric stability. One of the models, the following equation, should only be used if no measure of turbulence intensity is available:

$$(v_j)_t^{LT} = \beta_1 (v_j)_r^{LT} + \beta_5 (v_j^2)_r^{LT} + \alpha + \varepsilon_j \tag{58}$$

where the parameters β_1, β_5 and α depend on the reference site wind direction sector, θ_r. β_1 is the linear coefficient, β_5 the quadratic coefficient, α the offset and ε is white noise.

It is proposed that the other model, Eq. (59), be used when data are available for the temperature gradient or difference, ΔT_r, between the temperatures measured at two heights at the reference site:

$$(v_j)_t^{LT} = \beta_1 (v_j)_r^{LT} + \beta_4 (v_j)_r^{LT} (\Delta T_j)_r^{LT} + \alpha + \varepsilon_j \tag{59}$$

In Eq. (59), β_1 is the linear coefficient, β_4 the coefficient of interaction, α the offset and ε is white noise.

Thøgersen et al. [105] report that WindPro software has implemented second-order models with non-zero offset:

$$(v_j)_t^{LT} = \beta_1 (v_j)_r^{LT} + \beta_5 (v_j^2)_r^{LT} + \varepsilon_j \tag{60}$$

The regression parameters are calculated with the least squares method, using an optimisation algorithm called Amoeba, which is described by Press et al. [75].

Riedel et al. [107] also proposed using a second-order polynomial with zero offset, Eq. (60), in each of the dynamically calculated direction sectors. To determine the parameters β_1 and β_5, the authors proposed a novel strategy. The algorithm used attempts to minimise the weighted sum of the squares of the differences between the histograms of the measured and estimated short-term target site wind speeds using a chi-square test.

MINT [28] which operates with statistical and mathematical software packages developed for scientific and commercial ends, can perform regressions using first-, second- and third-order polynomials, Eq. (61). That is, by using a linear, quadratic or cubic fit. If the method has to include a zero intercept, this software also allows to select this option. The parameters of the models can be estimated by means of ordinary linear, orthogonal and quantile regression:

$$(v_j)_t^{LT} = \beta_1 (v_j)_r^{LT} + \beta_5 (v_j^2)_r^{LT} + \beta_6 (v_j^3)_r^{LT} + \alpha + \varepsilon_j \tag{61}$$

McKenzie et al. [130], in addition to second- and fifth-order polynomials, considered the use of second- and third-order Taylor polynomials.[27] Rebbeck [55] analysed models based on fifth-order polynomials and cubic splines. The author reached the conclusion that none of the models operated substantially better than a conventional linear regression model of two parameters.

4.4.4. Non-linear methods

Riedel et al. [107], with respect to the quadratic term of Eq. (60), assume that it will be low in relation to the linear term. So, they tried to bound the quadratic coefficient, β_5, of Eq. (60), by taking a cut-off wind speed, $v_{cut-off}$, off 10 m/s and a maximum wind speed, $v_{\max}$, of 30 m/s. So, Eq. (60) is replaced with the following equation, with the parameters now being β_1 and η:

$$(v_j)_t^{LT} = \beta_1 (v_j)_r^{LT} + \frac{v_{cut-off} \sin(\eta)}{v_{\max}^2} (v_j^2)_r^{LT} + \varepsilon_j \tag{62}$$

According to Derrick [17], experience shows that though the relationships between the wind speeds are not always linear, in most cases they seem to fit well. However, if a linear fit is not suitable a power regression can be employed using a function of the type indicated in the following equation, with parameters η and δ:

$$(v_j)_t^{ST} = \eta [(v_j)_r^{ST}]^{\delta} \tag{63}$$

In this case, the linear regression is performed after taking logarithms on both sides of the following equation:

$$\log[(v_j)_t^{ST}] = \log(\eta) + \delta \log[(v_j)_r^{ST}] \tag{64}$$

King and Hurley [98] also evaluated the fit of a power law curve, Eq. (63), but despite their optimism it did not provide any significant improvement over the first-order linear regression methods, Eq. (21). The method of fit they employed was the one commonly used for fitting power curves to data. That is, they took logarithms to both data series and then applied a linear regression of least squares (on the line that starts at the origin and passes through the centroid). However, as it takes logarithms, this method gives more weight to the set of lowest wind speeds, which is not desirable. The authors say they are open to suggestions as to other methods of fit of the power curve that might give better results.

McKenzie et al. [130], Clive [133] and van Lieshout [135], for the development of the methodology they propose, take as their starting point the hypothesis that wind regimes at the target

[27] It should be mentioned that the authors do not strictly use the Taylor Expansion Series as a regression fit, as they estimate the parameters beforehand through the fit of a Weibull distribution to the reference and target site data.

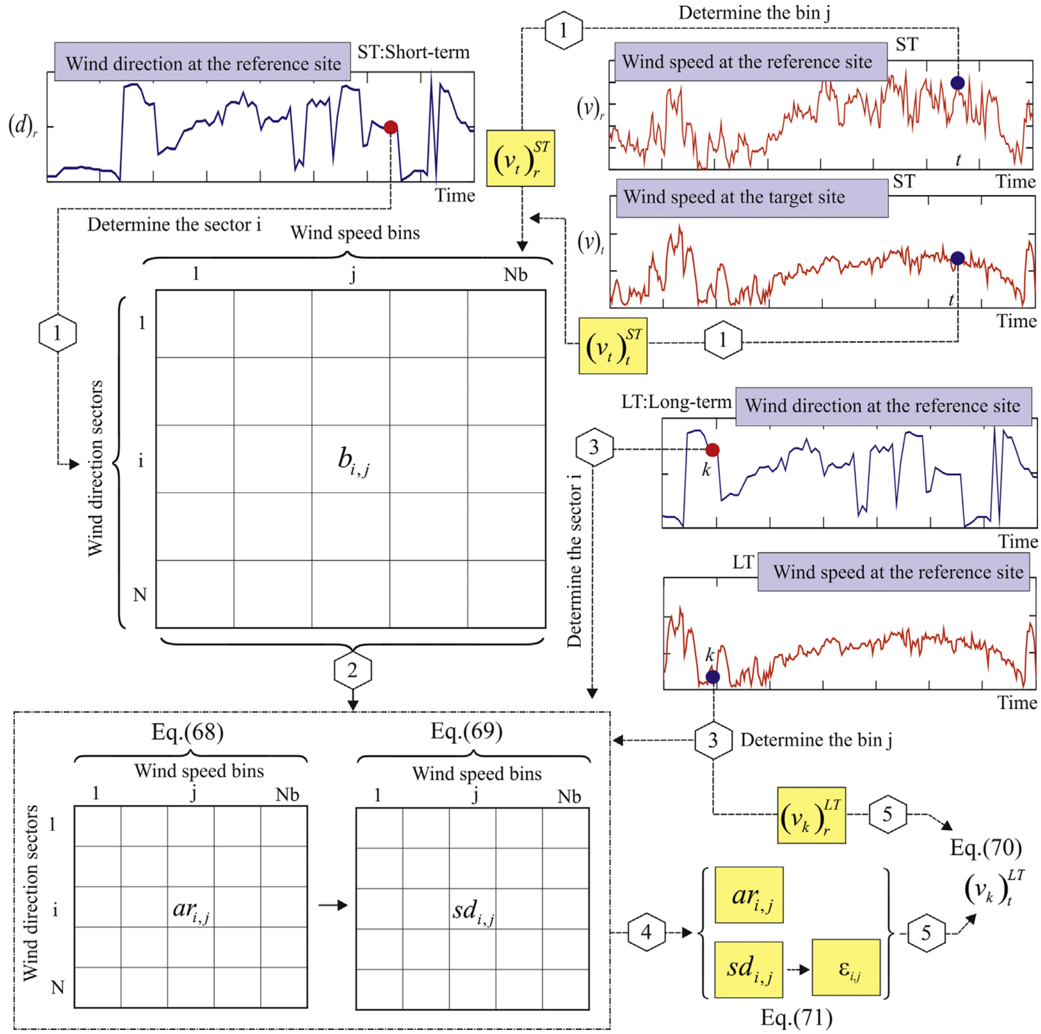

Fig. 16. Flow diagram created by the authors to facilitate interpretation of the block of tasks of the algorithm followed by the discrete method of Mortimer [136].

and reference sites can be represented by Weibull distributions [144,164].[28] Likewise, they consider that if a monotonic relationship exists between the target and reference site wind speeds, then the values of their cumulative distribution functions (CDF) will be equal. The three parameter Weibull distribution function (K, shape parameter; C, scale parameter and γ, location parameter) can be expressed in closed form and is susceptible to linearisation. So, making the cumulative distribution function of the target site equal to the cumulative distribution function of the reference site leads to the expression shown in the following equation:

$$\left[\frac{(v_j)_r^{ST}-\gamma_r^{ST}}{C_r^{ST}}\right]^{K_r^{ST}} = \left[\frac{(v_j)_t^{ST}-\gamma_t^{ST}}{C_t^{ST}}\right]^{K_t^{ST}} \Rightarrow (v_j)_t^{ST} = C_t^{ST}\left[\frac{(v_j)_r^{ST}-\gamma_r^{ST}}{C_r^{ST}}\right]^{K_r^{ST}/K_t^{ST}} + \gamma_t^{ST} \tag{65}$$

From Eq. (65) the authors [130,133] construct Eq. (66), considering that $\gamma_r^{ST}=0$ and where the parameters β and α are given by Eq. (67):

$$(v_j)_t^{LT} = \beta (v_j^{\delta})_r^{LT} + \alpha \tag{66}$$

[28] It should be mentioned that, though the two parameter Weibull distribution has been widely used in renewable energy related literature to represent wind regimes [2,144,164], in several references [144,165–167] sites have been indicated where the wind data do not obey a two parameter Weibull distribution but rather other probability density functions.

$$\delta = \frac{K_r^{ST}}{K_t^{ST}};\ \beta = C_t^{ST}(C_r^{ST})^{-\delta};\ \alpha = \gamma_t^{ST} \tag{67}$$

Only in the event that the shape parameters of both distributions are equal ($\delta=1$), will Eq. (66) be a first-order linear function. So, as stated by Clive [133], an error is being made if, in linear regression based MCP methods, the shape parameters of the Weibull distribution are different at the reference and target sites.

Sreevalsan et al. [134] present and discuss an MCP method based on fast Fourier transforms (FFT), using as starting point a previous study undertaken by Hunt and Nason [168]. Considering the difference in amplitude for three months of time series data of the target and reference sites, the method generates a one year time series of the target site wind speeds. According to the authors, the comparison made between the simulated and measured wind data showed a good correlation. However, they state that it remains to be seen whether this procedure can be adopted for data synthesis of multiple years with small deviations. They also comment that their study is only a preliminary one and that more data series from other sites need to be analysed before general validation of the results.

According to Hunt and Nason [168], wavelet methods have been shown to provide more reliable estimations than prediction methods based on a simple linear regression. Moreover, they report that their models have the added benefit of often being physically interpretable. However, the computational complexity

of the model is very high when, for example, one year's worth of data is used.

4.4.5. Probabilistic methods

Various references have proposed the use of MCP methods based on probabilistic methods [125,130,133,136–143].

Mortimer [136] proposes a method he calls the discrete method, in which, in the concurrent data period, the data of both sites are distributed in a matrix **B**, in which an element, $b_{i,j}$, represents the wind speed bin i (with bin width of 1 m/s) and wind direction sector j of the reference site, Fig. 16. So, this method is based on the hypothesis that the target and reference site wind directions coincide. Each wind speed bin i of the reference site wind direction bin j contains the set of ratios calculated between each target site wind speed, $(v_k)_t^{ST}$, which is paired by date and time k with the corresponding reference site wind speed, $(v_k)_r^{ST}$. Two matrices are generated from the matrix **B**: one, **AR**, in which each element, $ar_{i,j}$, represents the average ratio corresponding to wind speed bin i of the reference site wind direction sector j, Eq. (68), and another, **SD**, in which each element, $sd_{i,j}$, represents the standard deviation of the ratios of wind speed bin i and wind direction sector j of the reference site, Eq. (69):

$$ar_{i,j} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} (b_{i,j})_k \tag{68}$$

$$sd_{i,j} = \frac{1}{n_{ij}-1} \sum_{k=1}^{n_{ij}} [(b_{i,j})_k - ar_{i,j}]^2 \tag{69}$$

where $n_{i,j}$ is the number of ratios stored in the element $b_{i,j}$ of matrix **B**.

Using the **AR** matrix of the means of the ratios and the **SD** matrix of the standard deviations of the ratios, the n wind speeds of the long-term target site wind speed series are estimated as a function of the n data of the long-term reference site wind speed $(v_k)_r^{LT}$ and direction $(d_k)_r^{LT}$ series, through the following equation:

$$(v_k)_t^{LT} = [ar_{i,j} + \varepsilon_{i,j}](v_k)_r^{LT};\ (v_k)_r^{LT}, (d_k)_r^{LT} \in b_{i,j};\ k = 1 \ldots n \tag{70}$$

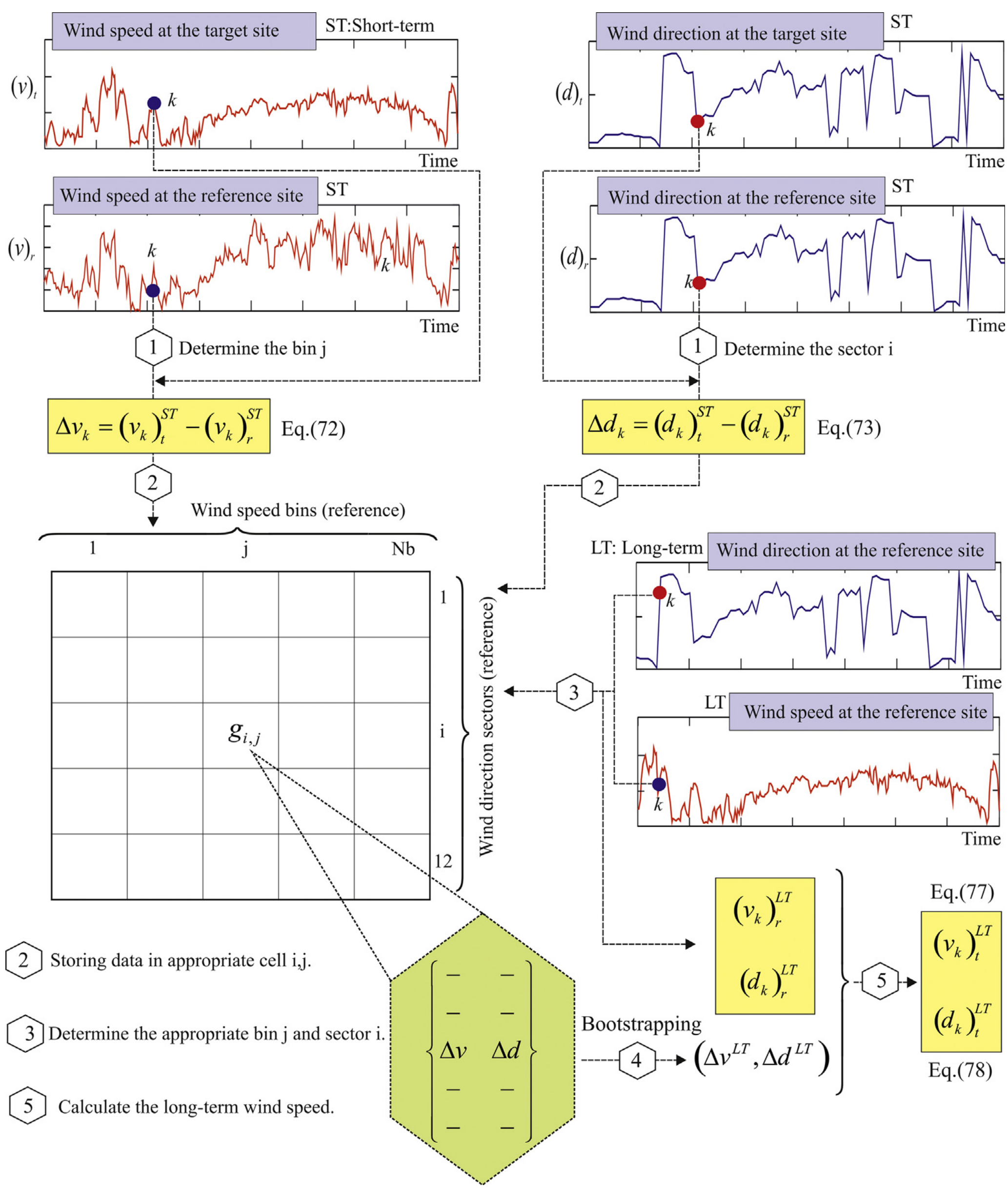

Fig. 17. Flow diagram created by the authors to facilitate interpretation of the block of tasks of the algorithm followed by the matrix method [24,105] when using the probability distribution associated with the variables Δv and Δd measured during the concurrent data period.

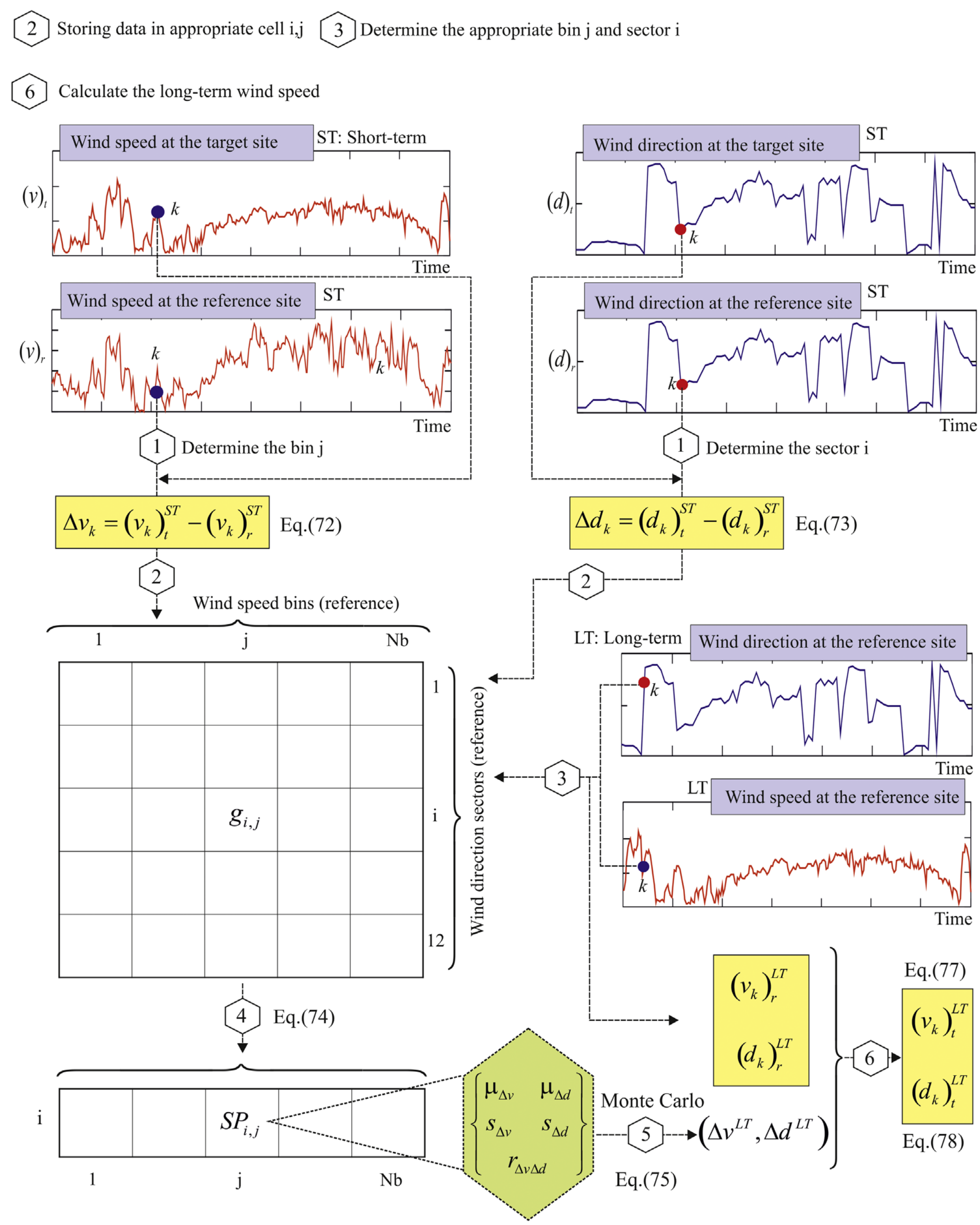

Fig. 18. Flow diagram created by the authors to facilitate interpretation of the block of tasks of the algorithm followed by the matrix method [24,105] when using polynomial functions together with a bivariate Gaussian distribution.

where $\varepsilon_{i,j}$ is a random variable with symmetric triangular distribution [169][29] used with the idea of taking into consideration the corresponding standard deviation, $sd_{i,j}$, of the ratios or their variability. Given a random value U extracted from a uniform distribution in the interval (0, 1), then the random variable, $\varepsilon_{i,j}$, is given by the following equation:

$$\begin{aligned} &0 \le U < 1/2 \ \Rightarrow \ \varepsilon_{i,j} = (2\sqrt{3U} - \sqrt{6})sd_{i,j} \\ &1/2 \le U \le 1 \ \Rightarrow \ \varepsilon_{i,j} = (\sqrt{6} - 2\sqrt{3(1-U)})sd_{i,j} \end{aligned} \tag{71}$$

Mortimer [136], as a result of tests carried out, concludes that the discrete method can predict extreme wind speeds better than first-order linear regression methods.

[29] The triangular distribution is used to describe a variable which is limited to a particular range but which very probably has values in the middle of that range.

Sheppard [56] modified Mortimer's method [136] by replacing the triangular distribution with a normal distribution. The direction sectors used by Sheppard [56] were one of the following: 22.5°, 45°, 90°, or 180°. In the procedure, the smallest possible wind direction sector size is chosen with the restriction that there are at least four observations in each bin.

Rogers et al. [106] used the method proposed by Mortimer [136] for purposes of comparison with three other MCP methods. In the algorithm that was implemented, the following actions were programmed: (a) if the reference site wind speed is below 1 m/s, the wind speed ratio takes the value 1, (b) if insufficient data for a particular element, $b_{i,j}$, of the matrix **B** are available to calculate an average ratio, then the ratio is determined as the ratio of the mean of all the wind speeds of the direction sector j and the standard deviation is taken as zero, (c) if Eq. (67) predicts negative long-term target site wind speeds, these wind speeds take zero value.

According to Thøgersen et al. [24,105], one of the MCP methods available in WindPro, called the matrix method, calculates in the concurrent data period the differences in wind speed (speed-up), Eq. (72), and the differences in wind direction (wind veer), Eq. (73), of the target and reference sites, sorting the results by reference site wind speed, $(v_k)_r^{ST}$, and wind direction, $(d_k)_r^{ST}$, Figs. 17 and 18. Results are stored in a matrix **G**, where each element or cell, $g_{i,j}$, corresponds to a wind speed bin, b_i, (with default width of 1 m/s) and a wind direction sector,$(\theta_j)_r$, (default value of 30°) of the reference site:

$$\Delta v_k = (v_k)_t^{ST} - (v_k)_r^{ST} \tag{72}$$

$$\Delta d_k = (d_k)_t^{ST} - (d_k)_r^{ST} \tag{73}$$

Since some elements, $g_{i,j}$, of the matrix might end up empty, the algorithm implemented can interpolate or extrapolate statistical properties from the other elements or cells with a sufficient number of data in order to fill in these gaps using for this purpose polynomial fits, Eq. (74). The properties estimated in each of the empty cells of the matrix are the means ($\mu_{\Delta v}$,$\mu_{\Delta d}$) and standard deviations ($s_{\Delta v}$,$s_{\Delta d}$) of the variables, Δv and Δd, as well as the correlation coefficient, $r_{\Delta v \Delta d}$, between them:

$$SP_{i,j} = \alpha + \sum_{kk=1}^{NP} \beta_{kk} [(v_j)_r^{ST}]^{kk} + \varepsilon_i \tag{74}$$

In Eq. (74), $SP_{i,j}$ is the estimated statistical property in cell $g_{i,j}$, the parameters α and β of the models depend on the wind direction sector i under consideration and on the estimated statistical property and ε_i is the residual term [2]. The implemented algorithm allows to select the order of the polynomial (NP), and this order can be different for each statistical property. NP=1 is the predetermined option for estimating the mean wind speed-up. However, to estimate the mean wind veer, the predetermined option considers that these changes are independent of wind speed and therefore proposes a zero-order polynomial. For standard deviations, the recommendation is to use NP=1 for wind direction changes and NP=2 for wind speed changes.

To estimate the long-term target site wind speeds, $(v_k)_t^{LT}$, the implemented algorithm admits two options. One of them, Fig. 17, consists of estimating the difference in wind speeds and the difference in wind directions as a function of the reference site wind speed and direction through the use of the probability distribution associated with the variables, Δv and Δd, measured during the concurrent period. This option is the one recommended if the number of samples in the relevant matrix cell is at least five. The other possibility, Fig. 18, lies in using the polynomial functions, Eq. (74), together with a bivariate Gaussian distribution, Eq. (75) [71], of the speed-up and wind veer:

$$Pdf(\Delta v, \Delta d) = A\exp \times \left[-\frac{((\Delta v - \mu_{\Delta v})^2 / s_{\Delta v}^2) - 2r_{\Delta v \Delta d}((\Delta v - \mu_{\Delta v})(\Delta d - \mu_{\Delta d}) / s_{\Delta v} s_{\Delta d}) + ((\Delta d - \mu_{\Delta d})^2 / s_{\Delta d}^2)}{2\sqrt{1 - r_{\Delta v \Delta d}^2}} \right] \tag{75}$$

where A is given by the following equation:

$$A = \frac{1}{2\pi s_{\Delta v} s_{\Delta d} \sqrt{1 - r_{\Delta v \Delta d}}} \tag{76}$$

Starting with the long-term reference site data time series, the long-term target site wind speed and direction time series is estimated through the following procedure. First, each pair of data of the series $[(v_k)_r^{LT}, (d_k)_r^{LT}]$ indicates the cell, $g_{i,j}$, of the matrix **G** that has to be used, Figs. 17 and 18. Second, if the option is chosen of using the data measured during the concurrent period (Fig. 17), then from the data contained in the above indicated cell, $g_{i,j}$, a pair of values $(\Delta v^{LT}, \Delta d^{LT})$ are selected by bootstrapping [170], which are considered to be long-term representative. If the second option is selected (Fig. 18), then the corresponding pair of values $(\Delta v^{LT}, \Delta d^{LT})$ are randomly generated from the bivariate Gaussian distribution associated with the above-indicated cell, $g_{i,j}$, through the Monte Carlo simulation method [147]. Subsequently, the long-term target site wind speed and direction are estimated using Eqs. (77) and (78), respectively.

$$(v_k)_t^{LT} = \Delta v^{LT} + (v_k)_r^{LT} \tag{77}$$

$$(d_k)_t^{LT} = \Delta d^{LT} + (d_k)_r^{LT} \tag{78}$$

Lambert and Grue [142] report that common matrix methods often perform well when compared with other methods, but they are normally used to generate frequency distributions rather than data time series. So, the authors propose a method they call the Matrix Time Series method (MTS), the algorithm of which is available in the MCP module of the WindoGrapher software [25] and which consists of an adaptation of the classic matrix method to enable estimation of long-term target site data time series. The procedure involves four basic steps. First, the reference and target site wind speeds of the concurrent data period are distributed in bins defined by the reference site wind direction and a matrix **U** is constructed for each reference site wind direction sector. Each matrix **U** constitutes a joint probability distribution representing each sector and period. Each cell, $u_{i,j}$, of each matrix **U** contains the number of times that, during the concurrent data period, target site wind speeds of bin i and reference site wind speeds of bin j were recorded. So, in each column j of a matrix **U** the probability mass function, PMF_j, is defined of the target site wind speed conditioned on the reference site wind speed of bin j. From each PMF_j it is possible to construct the cumulative probability mass function, $CPMF_j$, of the target site wind speed for the reference site bin j.

In the second step, a time series of percentiles[30] is constructed for the target site. For this, each pair of data, $(v_k)_r^{ST}$ and $(d_k)_r^{ST}$, of the reference site time series of wind speeds and directions allows to select the appropriate direction sector and appropriate bin j (column) of the matrix **U** of that sector, while the corresponding data, $(v_k)_t^{ST}$, of the target site time series of wind speeds indicates the corresponding percentile in the $CPMF_j$, Fig. 19.

For the third step, the algorithm synthesises a long-term time series of percentiles for the target site. For this, a first-order transition probability matrix of a Markov chain is constructed [171],[31] which stores the frequency with which the percentiles change from one state to another. By generating random numbers, artificial scenarios are created to fill in the missing data period of the target site time series. If the final synthetic percentile (at the end of the unknown time series) does not adequately coincide with the value of the first percentile of the measured time series, then the algorithm discards the simulated scenario and proceeds to construct others until a reasonable match is obtained between the final datum of the synthesised series and the first datum of the measured series. The algorithm has the option of smoothing out the time series of percentiles between the second and third steps using the technique of moving averages [67]. The authors [142] have verified that this option improves the accuracy of the results in many cases.

The fourth and final step attempts to transform the time series of synthetic percentiles into synthetic target site wind speeds, Fig. 20. Essentially, the procedure consists of reversing the second

[30] They are the 99 values that split the distribution into 100 equal parts. The percentile is a measure of non-central position that tells us how a value is positioned with respect to all the values of the sample.

[31] First- and second-order Markov chains have been used in renewable energy related literature to model wind data [172–173].

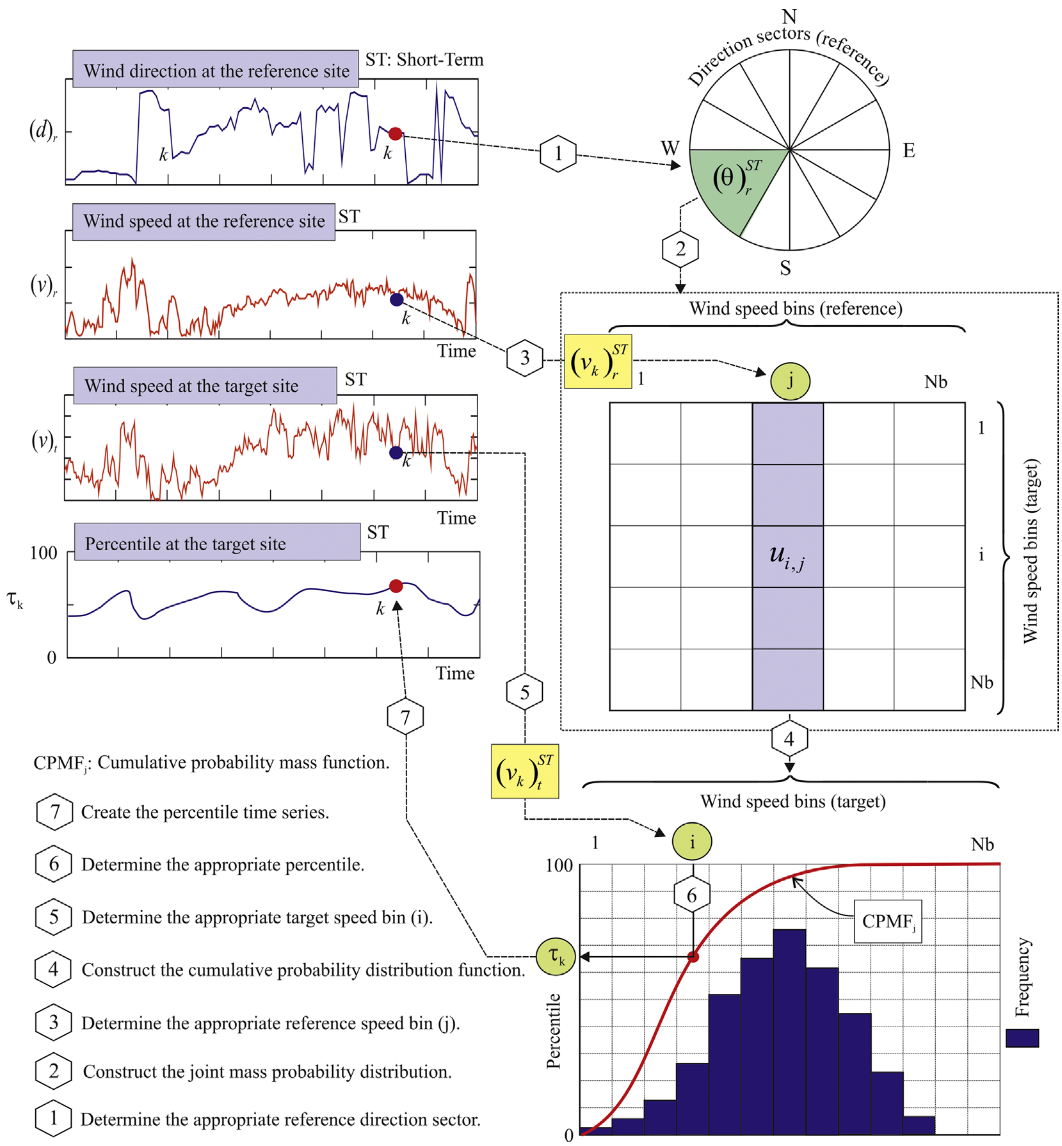

Fig. 19. Flow diagram created by the authors to show the first tasks of the algorithm followed by the MST method.

step. In this case, each pair of data, $(v_k)_r^{LT}$ and $(d_k)_r^{LT}$, of the long-term reference site time series of wind speed and direction allows to select the appropriate direction sector and, in that sector, the appropriate associated column (bin j) of the matrix **U**. The synthetic percentile k allows to calculate in the $CPMF_j$ of the bin j of that matrix the synthetic wind speed, $(v_k)_t^{LT}$, that corresponds to the target site data k.

Salmon and Walmsley [137] slightly modified a method first developed by Walmsley and Bagg [138], which allows that the reference site wind rose can be different to the target site wind rose. Initially, in the proposed methodology, a matrix **H** (3072×3072) is created which constitutes a joint relative frequency distribution of short-term wind speed, wind direction and atmospheric stability. That is, each element, $h_{i,j}^{ST}$, of the matrix represents the short-term relative frequency of occurrence of category i at the reference site and of category j at the target site. The number of categories, Nh, is obtained after classifying the short-term data of both stations by wind direction (16 equal sized sectors), wind speed (32 bins) and atmospheric stability (six bins).

The frequencies of occurrence of category i at the reference site and of category j at the target site are estimated from the following equation:

$$(F_i)_r^{ST} = \sum_{j=1}^{Nh} h_{i,j}^{ST};\ (F_j)_t^{ST} = \sum_{i=1}^{Nh} h_{i,j}^{ST} \tag{79}$$

The method calculates a matrix **C** of coefficients $c_{i,j}$, Eq. (80), which, assuming they remain constant over the long-term, are used to estimate the long-term target site relative frequencies, Eq. (81):

$$c_{i,j} = \frac{h_{i,j}^{ST}}{(F_i)_r^{ST}} \tag{80}$$

$$(F_j)_t^{LT} = \sum_{i=1}^{N} c_{i,j}(F_i)_r^{LT} \tag{81}$$

García-Rojo [139] also uses a procedure based on calculation of the joint probability mass function $p_{t-r}^{ST}(v_t, d_t, v_r, d_r)$ of the short-term wind speed (v) and direction (d) of the target and reference sites. Starting with this short-term joint probability mass function and the probability mass function of long-term wind speed and direction at the reference site, $p_r^{LT}(v,d)$, the probability mass

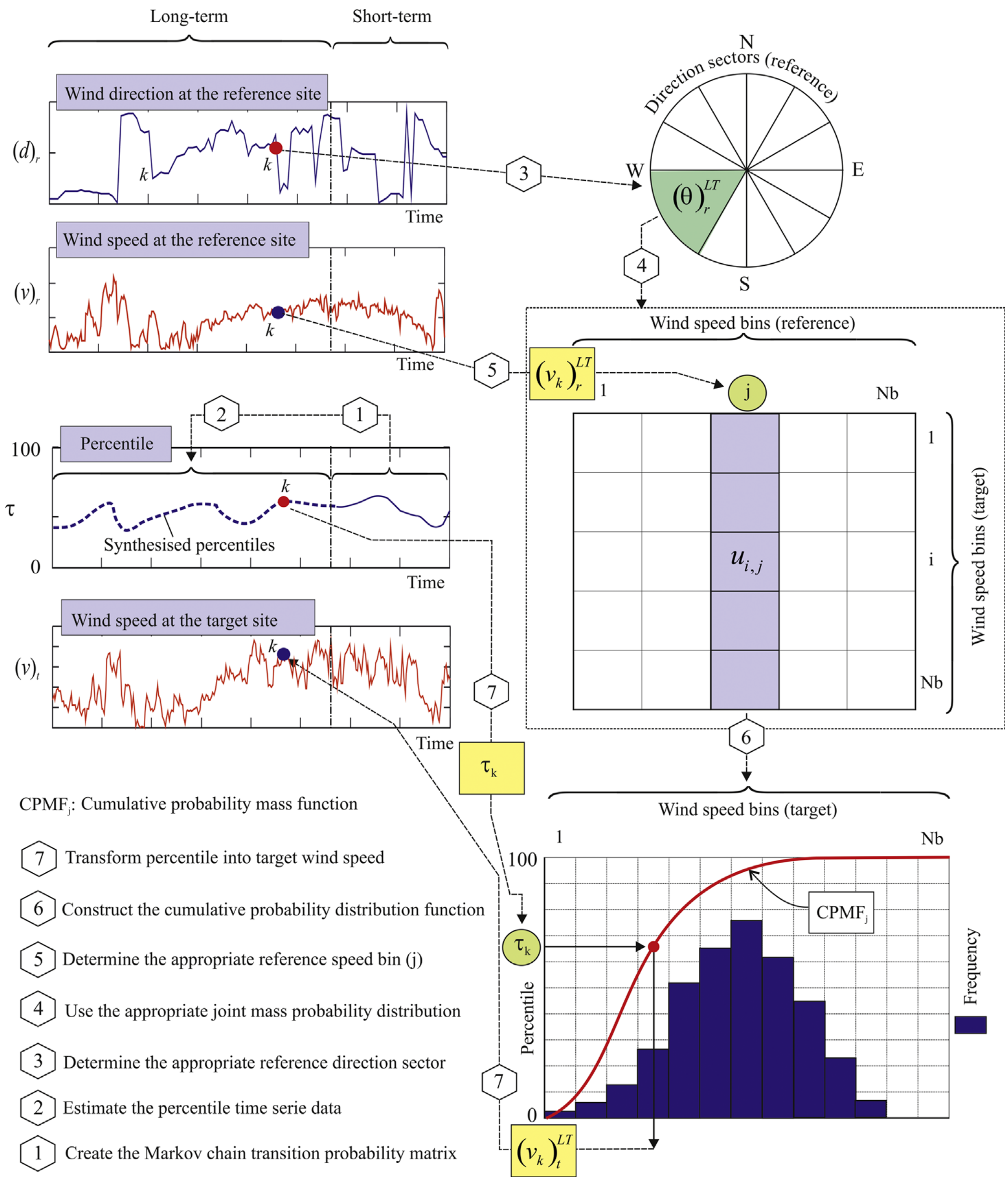

Fig. 20. Flow diagram created by the authors to show the last tasks of the algorithm followed by the MST method.

function of long-term wind speed and direction at the target site, $p_t^{LT}(v,d)$, is estimated:

$$p_t^{LT}(v_i,d_j)=\sum_{k=1}^{Nw}\sum_{z=1}^{Nd}p_{t-r}^{ST}(v_i,d_j,v_k,d_z)p_r^{LT}(v_k,d_z) \tag{82}$$

Nw and *Nd* are the numbers of bins of wind speed and direction, respectively, of the joint probability mass function. In this study, twelve 30° bins of wind direction were used to define the target and reference site wind direction. The widths of the wind speed bins were 1 m/s.

Carta and Velázquez [141] proposed a new MCP method, the algorithm of which uses the target site wind speed probability distribution function conditioned on the reference site wind speed. Contingency-type bivariate distributions with specified marginal distributions were used for this purpose:

$$f_t^{ST}(v_t|v_r)=\frac{1}{2\sqrt{S2}}f_t^{ST}(v_t)\left\{\frac{(\psi-1)[S1-2\psi F_t^{ST}(v_t)][S1-2\psi F_r^{ST}(v_r)]}{S2}+\psi+1\right\} \tag{83}$$

where $S1$ is given by Eq. (84) and $S2$ is given by Eq. (85):

$$S1=1+[F_t^{ST}(v_t)+F_r^{ST}(v_r)](\psi-1) \tag{84}$$

$$S2=S1^2-4\psi(\psi-1)F_t^{ST}(v_t)F_r^{ST}(v_r) \tag{85}$$

In the above equations, $F_t^{ST}(v_t)$ and $F_r^{ST}(v_r)$ are the cumulative distribution functions of the wind speed at the target and reference sites, respectively. $f_t^{ST}(v_t)$ is the wind speed probability density function of the target site. ψ is the coefficient of association which needs to be estimated.[32] To estimate ψ, the authors equate the values of the correlation coefficient of the sample and the model.

The marginal distributions used were two-component mixture Weibull probability density functions [144,165–166], which depend on five parameters.

[32] This bivariate distribution has only one parameter, in addition to the marginals.

In the proposed method, from the short-term conditional probability density function, Eq. (83), and the long-term reference site wind speed probability density function, $f_r^{LT}(v)$, the long-term target site wind speed probability density function, $f_t^{LT}(v)$, is estimated:

$$f_t^{LT}(v_t) = \frac{1}{CNF}\int_0^\infty f_t^{ST}(v_t|v_r) f_r^{LT}(v_r)dv_r \tag{86}$$

where CNF is a normalisation factor:

$$CNF = \int_0^\infty \int_0^\infty f_t^{ST}(v_t|v_r) f_r^{LT}(v_r)dv_r\, dv_t \tag{87}$$

The stratified k-fold cross-validation method with k partitions was used to train and test the model [174]. Using data series recorded in the Canary Islands (Spain), the authors concluded that this method gave better results in most cases than those obtained with the variance ratio method [106], Weibull scale method [130,133] and a modification of the method developed in [139], which they used for comparison purposes.

The modification of the method described in [139] consisted of using the probability density function of wind speed and direction at the target site conditioned on the wind speed and direction at the reference site, $p_t^{ST}(v_i, d_j|v_k, d_z)$, Eq. (88), instead of the joint probability mass function of wind speed and direction, $p_{t-r}^{ST}(v_t, d_t, v_r, d_r)$[33]:

$$p_t^{LT}(v_i, d_j) = \frac{1}{CNF}\sum_{k=1}^{Nw}\sum_{z=1}^{Nd} p_t^{ST}(v_i, d_j|v_k, d_z) p_r^{LT}(v_k, d_z) \tag{88}$$

where CNF is a normalisation factor:

$$CNF = \sum_{i=1}^{Nw}\sum_{j=1}^{Nd}\sum_{k=1}^{Nw}\sum_{z=1}^{Nd} p_t^{ST}(v_i, d_j|v_k, d_z) p_r^{LT}(v_k, d_z) \tag{89}$$

Romo et al. [125] also propose models based on conditional probability density functions. They have presented two models, one based on a bivariate normal distribution [143,175] and the other based on a bivariate Weibull distribution function, Eq. (90), whose marginal distributions are two-parameter Weibull distributions:

$$F(v_t, v_r) = 1 - \exp\left\{-\left[\left(\frac{v_t}{K_t^{ST}}\right)^{C_t^{ST}/\lambda} + \left(\frac{v_r}{K_r^{ST}}\right)^{C_r^{ST}/\lambda}\right]^{\lambda}\right\} \quad 0<\lambda\le 1 \tag{90}$$

where λ is a parameter which controls the degree of association between the two variables. The five parameters of the bivariate Weibull distribution are estimated using the maximum likelihood method.

In their work, the authors generate a time series of synthetic data with a controlled correlation structure and with the desired parameters of the distribution laws to evaluate the different MCP methods under consideration. Among their conclusions, they report that only one of the five models tested that based on the bivariate Weibull conditional probability function was able to accurately predict all five of the performance metrics under consideration.

Some authors [130,133] have proposed the use of an empirical method known as the Weibull scale method, based on parameter relationships of the two-parameter Weibull distribution law. This method takes as its starting point the assumption that the reference and target site wind speeds follow a two-parameter Weibull distribution. Additionally, the existence is considered of a linear relationship between, on the one hand, the scale parameters (C), and, on the other, the shape parameters (K) of the probability density functions of the target and reference sites, Eq. (91). So, essentially it is more a linear than probabilistic method:

$$\frac{\lambda_t^{ST}}{\lambda_r^{ST}} = \frac{\lambda_t^{LT}}{\lambda_r^{LT}} \quad \Rightarrow \quad \lambda_t^{LT} = \frac{\lambda_t^{ST}}{\lambda_r^{ST}}\lambda_r^{LT}; \tag{91}$$

In Eq. (91), λ represents the scale or shape parameter under consideration.

This method has been implemented in various wind industry software applications [24,25] and the Weibull parameters are calculated for each of the direction sectors in which the reference and target site wind data are distributed.

According to Thøgersen et al. [24,105], one advantage of this method is its ability to adapt to the nature of the wind at most sites, but care should be taken when applying this method at sites where the Weibull distributions are not representative of their wind regimes. When considering frequencies, the modified long-term distribution must be normalized for the N direction sectors, θ_r, under consideration:

$$f_t^{LT}(\theta_r) = \left\{\frac{f_t^{ST}(\theta_r)}{f_r^{ST}(\theta_r)} f_r^{LT}(\theta_r)\right\}\left\{\sum_{\theta=1}^{N}\frac{f_t^{ST}(\theta_r)}{f_r^{ST}(\theta_r)} f_r^{LT}(\theta_r)\right\}^{-1} \tag{92}$$

4.5. *Estimation methods of long-term wind characteristics using multiple reference stations*

Though most of the MCP methods used to date in the wind industry have employed a single reference station, methods have been proposed in the literature based on the use of various reference stations to estimate the wind characteristics at a target site.

Probst and Cárdenas [61] are of the opinion that it may often be advisable to consider various reference stations with simultaneously measured datasets to estimate the long-term wind characteristics of a target site. They state that a multiple linear regression is often assumed [66–67] between the wind speed variables. According to Brower [9], sometimes the different reference stations capture different aspects of the wind climate at the target site and a multiple linear regression can be a practical way of improving the general correlation, as the weights given to each reference station in the fit depend on the correlation of each station with the target site station and its statistical independence from the other stations. However, Brower [9] warns that there is a risk of obtaining poor results as a consequence of an over-specified fit if the number of stations used is too high and they are strongly correlated.

Walls et al. [176] undertook a study which, for estimation of the long-term target site wind speed, was based on information provided by two reference stations ($r1$, $r2$) located 18 km and 29 km from the target site (t), and on the use of what they term planar regression:

$$(v_i)_t^{LT} = \beta_1 (v_i)_{r1}^{LT} + \beta_2 (v_i)_{r2}^{LT} + \alpha + \varepsilon_i \tag{93}$$

The authors reported that the mean absolute error and the standard deviation of errors fell when planar regression was used compared to the results obtained when using only one of the two reference stations. However, planar regression did provide a small improvement in the accuracy of the estimation based on short-term data periods.

Oh et al. [177] propose a strategy, which they call the complementary MCP technique, to reconstruct wind data time series from stations with missing data periods that prevent having continuous long-term data series of sufficient length for proper analysis. The proposed strategy requires various nearby stations that can provide the information needed to fill in the missing data period/s. Their proposal does not involve regression models which simultaneously use information from various stations, but rather

[33] It should be mentioned that this fact was not correctly explained in [141].

simple linear regression techniques between two stations to restore the missing wind speeds and directions using the station with the higher coefficient of determination, R^2, as the reference station.

Casella [178], based on the argument that when various reference stations are available it may be preferable to perform a multiple-correlation analysis to improve the accuracy of the target site estimation, proposes a new probabilistic MCP method, named the J-tris method. It constitutes an extension of the method published by García-Rojo [139], but which in this case uses a joint probability function constructed from the simultaneous measurements recorded at three weather stations. Here, the data of the short-term period of the target station are corrected by the long-term data from the two reference stations.

In the 1980s, Barros and Estevan [116] proposed using a method, which they called the Multiple Climatic Reduction Technique (MCRT), to estimate the mean annual wind speed and its yearly variation at a target site using just two or three months' worth of wind speed measurements for that site. In the procedure followed by the authors, if $\mathbf{V}(L \times M)$ is a matrix whose columns j are the time series with L data from M weather stations, it is possible to represent the elements $v_{i,j}$ through the following equation:

$$v_{i,j} = \sum_{k=1}^{L} a_{i,k} w_{k,j}; \quad i = 1 \ldots L; \quad j = 1 \ldots M \tag{94}$$

where $a_{i,k}$ are the elements of an orthogonal matrix $\mathbf{A}$ $(L \times L)$ formed by the eigenvectors of the product of matrix $\mathbf{V}$ $(L \times M)$ and its transposed matrix $\mathbf{V}^{\mathrm{T}}$ $(M \times L)$, and $w_{k,j}$ are the elements of a matrix $\mathbf{W}$ $(L \times M)$ which correspond to projections of the data from each station onto each row of $\mathbf{A}$ $(L \times L)$.

Like all MCP methods, this one is based on the assumption that the wind speeds have some degree of spatial correlation and, therefore, there exists some sort of dependence. Based on this, it is assumed in the procedure that the elements, $v_{i,j}$, can be represented with reasonable accuracy through the following equation:

$$v_{i,j} = \sum_{k=1}^{Na} a_{i,k} w_{k,j}; \quad i = 1 \ldots L; \quad j = 1 \ldots M \tag{95}$$

where $Na < \text{minimum}\{L, M\}$ is the number of eigenvectors that must be chosen so that Eq. (95) conserves a significant percentage of the variance expressed by the trace of $\mathbf{V}\mathbf{V}^{\mathrm{T}}$.

If the data of the target station $(j = t)$ are incomplete and only a certain length is available of the time series $i = T1 \ldots T2$, with $T1 \geq 1$, $T2 \leq L$ and $T2 - T1 + 1 < Na$, the values, $v_{i,t}$, can be expressed by Eq. (94), even though they have not been used to calculate the elements of the matrix $\mathbf{A}$.

The method proposed is based on minimisation of the mean square errors, Eq. (96), which are produced between the observations made at the target station $(j = t)$ and their representation by means of empirical orthogonal functions, Eq. (95):

$$\sum_{i=T1}^{i=T2} \varepsilon_{i,t}^2 = \sum_{i=T1}^{i=T2} \left(v_{i,t} - \sum_{k=1}^{Na} a_{i,k} w_{k,t} \right)^2 \tag{96}$$

Minimisation of Eq. (96) leads to a linear system of Na equations and Na unknowns $(w_{k,t})$. Once the unknowns have been found, the unknown wind speeds of the target station can be estimated by applying the following equation:

$$v_{i,t} = \sum_{k=1}^{Na} a_{i,k} w_{k,t}; \quad i = 1 \ldots T1 - 1 \wedge i = T2 + 1 \ldots L \tag{97}$$

In 1988, Barros and Schmidt [179] described a method to find the minimum span of data required to extend a climatic data series to a specified longer period using the MCRT.

Zaphiropoulos et al. [180] propose a method for wind speed and direction estimation using only a short-term data period from a target site and long-term data records from various nearby weather stations. The technique is based on the use of statistical analysis and the inclusion of kriging methodology based on the works of Haslett and Raftery [181]. In [180], the authors expand upon an earlier work of theirs [182] by using 10-min averaged instead of day-averaged wind data as well as by estimating the wind speed in 16 directions.

Briefly, the method involves using a matrix $\mathbf{V}$ $(M \times L)$ whose rows i are the time series of M weather stations whose data L were recorded over a long-term period, except for one $(i = t)$ where only a short-term period of measured data is available $j = T1 \ldots T2$, with $T1 \geq 1$ and $T2 \leq L$. In the proposed method, the general least square estimator of the mean of the square roots of the target site wind speeds[34] is given by the following equation:

$$\bar{v}_t^{LT} = \frac{\mathbf{Q}_t^T [\mathbf{v}^{ST} - \mathbf{v}^{LT}]}{q_{t,t}} \tag{98}$$

where $\mathbf{v}^{ST} = (\bar{v}_1^{ST}, \ldots, \bar{v}_i^{ST}, \ldots \bar{v}_M^{ST})^T$ and $\mathbf{v}^{LT} = (\bar{v}_1^{LT}, \ldots, \bar{v}_i^{LT}, \ldots \bar{v}_M^{LT})^T$ are vectors whose elements are the means of the square roots of the short-term, Eq. (99), and long-term, Eq. (100),[35] wind speeds calculated with the observed data of each of the M stations considered. $\mathbf{Q}_t^T = (q_{1,t}, q_{2,t}, \ldots, q_{t,t}, \ldots q_{M,t})^T$ is the transpose of the column vector, $z = t$, of the inverse of the square matrix $\mathbf{R}$ $(M \times M)$, whose elements are defined by Eq. (101).[36] In Eq. (101), $dd_{k,z}$ and $r_{k,z}$ represent the distance and the correlation coefficient, respectively, between the station k and the station z. $0 \leq \eta \leq 1$ and $\gamma \geq 0$ are two parameters of the equation that represents the relationship between the data $r_{k,z}$ and $d_{k,z}$:

$$\bar{v}_i^{ST} = \frac{1}{T2 - T1 + 1} \sum_{j=T1}^{T2} v_{i,j}; \; i = 1 \ldots M \tag{99}$$

$$v_i^{LT} = \begin{cases} \frac{1}{L} \sum_{j=1}^{L} v_{i,j} & (i \neq t) \\ 0 & (i = t) \end{cases}; \; i = 1 \ldots M \tag{100}$$

$$r_{k,z} = \begin{cases} \eta \exp(-\gamma dd_{k,z}) & (k \neq z) \\ 1 & (k = z) \end{cases}; \; i = 1 \ldots M \tag{101}$$

$\bar{v}_{i=t}^{ST}$ is an estimator that does not take into consideration the data from the other sites, while $\bar{v}_t^{LT}$ is an estimator that does make use of these data. However, it should be pointed out that the technique used by the authors only gives an estimation of the overall speed measure for the target site. That is, the method does not offer predictions for the different periods in which the data were measured.

Eq. (98) is based on the assumption that Eq. (101) and the spatial covariance structure [180,182] hold, and on the assumption that no time dependence exists. Given that in this approach there are many underlying kriging procedures, $\bar{v}_t^{LT}$ is called the kriging estimator. The authors assume that $\bar{v}_t^{LT}$ follows a normal law and they formulate an expression for the variance of this estimator.

Denison et al. [184] propose a method for spatial data analysis with the aim of predicting the wind speed at a new site for which only a short-term period of data is available. The authors use a completely non-linear model, specifically the BMARS (Bayesian

[34] The distributions of the wind speeds at the stations analyzed by the authors were significantly asymmetric. So, they applied a square root transformation with the aim of stabilizing the variance and making the distribution approximately normal, as also performed in other studies [183].

[35] Except for the target site, where the long-term mean is here considered zero.

[36] The equation picks up the fact that the correlation between the data of the stations falls as the distance between these stations rises.

multivariate adaptive regression spline), given its ability to fit well to data in the presence of irrelevant predictors and because of its overall interpretability. In their study, they compare their method with that introduced by Haslett and Raftery [181], using a dataset recorded in areas of topographical complexity in the island of Crete.

Denison et al. [184] state that this method is particularly suitable for topographies of complex terrain where, unlike the data analysed by Haslett and Raftery [181], no systematic relationship exists between the distance between stations and the correlation. The authors stress that the novelty of the model that they propose lies in the fact that it can be used to give a good predictive performance even in the presence of a small spatial correlation structure.

In recent years, data mining and its applications have aroused much interest [185]. In this context, automatic learning techniques have been proposed that enable the use of two or more reference stations. These include techniques that employ statistical learning algorithms, such as Bayesian networks (BNs) [103], techniques based on biology-inspired algorithms, such as Artificial Neural Networks (ANNs) [102] as well as other alternative data mining techniques, including multivariate statistical techniques [186].

Carta et al. [103] propose using Bayesian network classifiers to estimate the long-term mean wind speed frequency distribution at a site with few wind resource measurements.

The method they propose allows to use multiple reference stations with a long history of wind speed and direction measurements. That is, the proposed method can use regional information of the wind resource. The intelligent system that is used, the knowledge base of which is a joint probability function of all the variables of the model, employs efficient calculation techniques of conditioned probabilities to carry out its reasoning. This allows to perform efficiently the automatic learning of the model and to perform the inference from the available evidence. The proposed model was applied to wind speeds and wind directions recorded at four weather stations located in the Canary Islands (Spain), Fig. 21. Among the conclusions reached by the authors of the study was that the BN with three reference stations gave lower errors between the real and the estimated long-term mean wind turbine energy output than those obtained with two standard MCP algorithms used for comparison purposes.

Artificial Neural Networks have been proposed and applied in renewable energy systems [187]. In 2000, Addison et al. [188] proposed a neural network version of the MCP algorithm for estimating wind energy yield. The authors investigated the possibility of using neural networks to make long-term energy yield predictions for a potential wind site. In their study, they considered the effectiveness of neural networks in the prediction of wind speed at a target site using wind speed and direction measurements from just one reference station. The authors compared this technique with standard MCP algorithms used in the wind energy industry. They concluded that with this technique an improved accuracy of 5–12% in the prediction was achievable. Similar proposals have subsequently been made in other studies [189].

A large number of the proposed models that use various reference stations use only the recorded wind speeds at these stations as the input layer signals. The topologies of the ANNs used have only a single neuron in the output layer, with the output signal being the estimated wind speed at the target site. In other words, most of the models used ignore the influence that the wind direction at the reference sites might have on the estimation of target site wind speeds [190–192]. López et al. [193] did study the influence of wind direction, in angular magnitude, on wind speed estimation in complex terrain. However, of the five anemometer stations used in the study, the same one was always used as the target station and the correlation coefficients between stations were of the same order. Velázquez et al. [102] presented the results of a study undertaken in the Canary Archipelago (Spain) in which the influence was investigated of various factors (number of reference stations, degree of correlation between target and reference station wind speeds, wind direction and the manner in which the direction signal was introduced into the input layer of an ANN) on the estimation of wind turbine generated power at a target site. The authors concluded that when wind direction was used in angular magnitude as input signal of the ANNs, Fig. 22, the metrics obtained gave better results than those obtained when that direction signal was omitted. They also report that estimation errors tend to fall as the number of reference stations used rises. However, on occasions the addition of some reference stations did

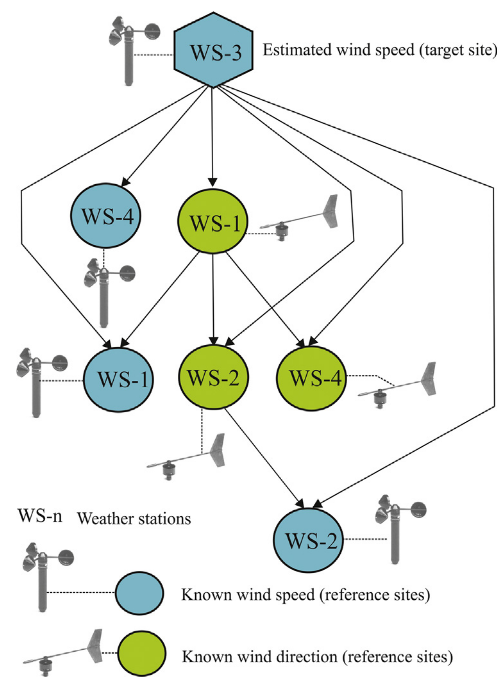

Fig. 21. Graph of a Bayesian network model described in Ref. [103]. The relationships of independence/dependence between the variables (wind speeds and directions) which make up the model can be seen in the graph. It is similarly possible to construct the graph of a model which allows to estimate the target site wind direction.

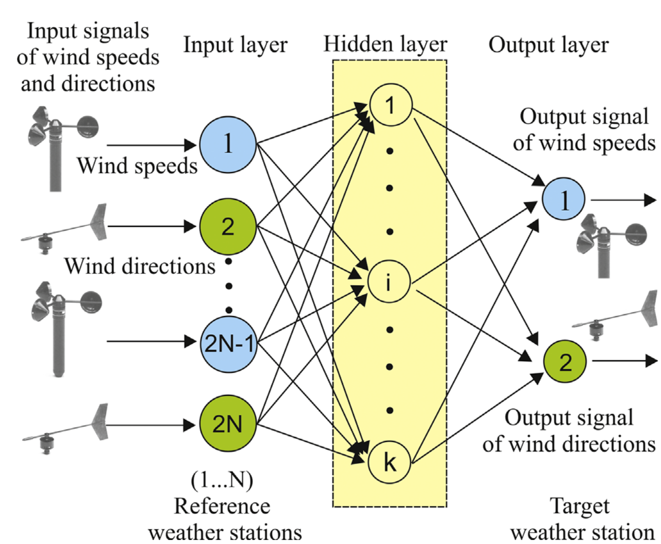

Fig. 22. Schematic diagram of an ANN with 2*N* wind speed and angular wind direction input signals of *N* reference stations and two wind data output signals of the target station.

not lead to an improvement in the estimation error. According to Faulkner [48], though some studies show that ANNs can give good results, these have not been much better than other methods and, in general, they are difficult to implement.

Patané et al. [194] state that data from anemometer stations with long measurement periods and numerical weather models can provide a large amount of potential inputs for MCP methods. To process all the available data, the authors developed a multivariate MCP methodology (MMCP). This technique efficiently takes into account all the inputs and extracts the maximum amount of relevant information at the target site. Given that the selected variables can be highly intercorrelated and that some of them may not provide information about the target site wind characteristics, the MMCP method is developed in a two-stage process. In the first stage, the input variables are analysed and those that contain little or redundant information about the target site wind characteristics are discarded. In the second stage, a multivariate regression is performed, similar to the analysis of principal components, based on the cross-correlation matrix of the selected variables and the time series of target site wind speeds [186]. As a result, an extrapolation of the wind speed time series at the target site is obtained. According to the authors [194], it can be concluded from the results of their tests that the MMCP method was more accurate than a standard MCP method used for purposes of comparison. Their method increased the quality of the estimation of the long-term wind resource by 19% with respect to the standard MCP method.

4.6. Estimation methods of the long-term wind direction

It can be seen from the descriptions of the methods of wind speed estimation given in the preceding sections that long-term target site wind direction is not commonly estimated in traditional MCP methods. Generally, the long-term target site wind direction is assumed to be the same as the wind direction recorded over a year long period at that site. However, several studies have been made which do propose specific procedures for estimating the long-term target site wind direction [24,98,105,107,122] and some of the technologies proposed for wind speed estimation also allow estimation of long-term wind direction [74,102–103,160].

Riedel et al. [107] propose using a functional relationship between the short-term measured wind directions at the reference and target sites to estimate the long-term target site wind direction. Their proposal involves eliminating data of the wind direction series that correspond to wind speeds below 3 m/s and using Chebyshev polynomials [195] as a function for predicting the wind direction. Similarly, Derrick [122] proposes fitting a polynomial to the wind direction data of the concurrent measuring period. Subsequently, this polynomial has to be applied to the long-term wind direction distribution of the reference station to obtain a prediction of the long-term target site wind direction distribution. Derrick [122] proposes elimination of data of the wind direction series below 4 m/s. According to the author, the 'meandering behaviour' of the vanes of the anemometers with low wind speeds may completely conceal the trends of wind directions for high wind speeds.

The SpeedSort, DynaSort and Scatter methods, proposed by King and Hurley [98], perform veer analysis for each data of the time series in the concurrent data period. Veer is defined as the difference between the wind direction at the target and reference sites. Average veer for a given direction sector is determined as the mean of all the hourly veer values in that sector. The sector-based veers are applied to the long-term reference site data to generate a wind direction frequency table for the target site. According to the authors, Scatter was the best method for estimating the long-term wind rose shapes.

The matrix method [24,105], as commented on earlier, also allows to estimate the long-term wind direction by using Eq. (78).

The methods proposed by Achberger [160] and Nielsel et al. [74], given that they estimate the components, $[(v_j)_x]_t^{LT}$ and $[(v_j)_y]_t^{LT}$, of the long-term wind speed at the target site, can also estimate the long-term wind direction through the following equation:

$$(d_j)_t^{LT} = \begin{cases} tan^{-1}\left\{\frac{[(v_j)_x]_t^{LT}}{[(v_j)_y]_t^{LT}}\right\}; & [(v_j)_x]_t^{LT} \geq 0, [(v_j)_y]_t^{LT} > 0 \\ \frac{\pi}{2}; & [(v_j)_x]_t^{LT} > 0\,, [(v_j)_y]_t^{LT} = 0 \\ \pi + tan^{-1}\left\{\frac{[(v_j)_x]_t^{LT}}{[(v_j)_y]_t^{LT}}\right\}; & [(v_j)_y]_t^{LT} < 0 \end{cases}$$

$$(d_j)_t^{LT} = \begin{cases} 2\pi + tan^{-1}\left\{\frac{[(v_j)_x]_t^{LT}}{[(v_j)_y]_t^{LT}}\right\}; & [(v_j)_x]_t^{LT} \leq 0 \\ \frac{3\pi}{2}; & [(v_j)_x]_t^{LT} < 0\,, [(v_j)_y]_t^{LT} = 0 \end{cases} \tag{102}$$

According to Roger et al. [106], as a conclusion of the studies they made, the so-called "vector" model [74] increases the bias associated with linear regressions, but it does predict relatively well the wind direction distribution, even in sites with significant wind direction deviations with respect to the reference site. However, they state that though the vector method seems to be suitable for wind direction modelling, more research is needed to find the best wind direction prediction method.

The techniques of Bayesian networks [103] and of ANNs [102] allow estimation of both long-term wind speed and direction.

5. Uncertainties in results of MCP analyses

As mentioned in Introduction section of this review, since the economic viability of a wind farm depends fundamentally on its energy production, tools like the MCP methods are of vital importance for estimating the long-term wind resource at wind farm candidate sites [2–9]. Nonetheless, wind resource estimation is only useful if the uncertainty associated with the estimation is properly understood by the agents involved. In other words by the investors, land owners, credit institutions, etc.

There are several sources of uncertainty [15,46,51–52] and among them is found the uncertainty associated with the estimation of long-term wind characteristics. To analyse the set of uncertainties, use has been made of deterministic and stochastic methods, such as the Monte Carlo method [46].

To assess the uncertainty of the predictions made, users of MCP methods only have the measured wind data of the concurrent period at both reference and target sites and the long-term data recorded at the reference site. This means that there is effectively little information available to determine the uncertainty. In addition, as pointed out by Bass et al. [104], the uncertainty related to the long-term predictions of MCP methods depends much more on the data than on the actual method used. That is, though some methods may turn out to be better than others at estimating the long-term wind characteristics of a particular target site, to a large degree the uncertainty depends on the extent to which the conditions imposed by these methods to ensure their effectiveness (described in Section 2) are met. In other words, uncertainty depends to a large extent on compliance with wind measurement protocols, on the similarity of the reference and target site climates, on the knowledge of seasonal and daily wind variations and on climate stability. Within the framework of these restrictions, various statistical models have been proposed to assess the uncertainty of the prediction of MCP methods. According to Faulkner [48], no robust or universal method has yet been found to quantify this uncertainty. However, there are some rough guides and rules of thumb that can be used. Some of these are provided in the studies that have been carried out on MCP methods

[47–48,50,104]. In general, they conclude that the longer the concurrent data period and the more extensive the long-term data time series for the reference site, the lower will be the uncertainty of results obtained with MCP methods. According to Faulkner [48], some studies have found a vague relationship between the correlation coefficient and uncertainty, but this relationship is neither consistent nor conclusive. Bass et al. [104] report that they detected a clear relationship between uncertainty and the differences in climate type or terrain complexity.

Faulkner [48] believes that a key question is whether the relationship that is established in the concurrent data period is representative of the long-term relationship. However, he also notes that there is no direct way of determining this and, therefore, there is no direct way of quantifying the uncertainty. Despite this, all the statistical models used to assess uncertainty have used measures of the variability in the relationship between the data of the reference and target sites during the concurrent data period to predict the variability of the predictions.

Derrick [17,122] uses the variances and covariances of the slope, β, and the offset, α, to estimate the standard deviation of the long-term wind speeds predicted with a linear regression model applied to different reference site wind direction sectors. However, according to Rogers et al. [47], this method underestimates the overall uncertainty and erroneously assumes serial independence as this is not the case for wind speeds. According to these authors, such an assumption may significantly affect the results and furthermore, they state, this approach can only be applied to MCP linear regression models.

There are various strategies that can be used to evaluate the level of accuracy of the models. One approach consists of using the full set of data from the concurrent period to train the model and assess its quality. However, this strategy favours those models which overfit the training dataset and do not generalise for other data. Rogers et al. [47] use the jackknife estimate of variance [170], while Carta et al. [103] and Carta and Velázquez [141] employ cross-validation [170]. In [103,141], the authors use 10-fold cross validation with the aim of using the variance of the 10 partial sampling errors to estimate the variability of the learning method with respect to the evidence.

6. Conclusions

This paper offers a general review of a wide collection of MCP methods proposed for use in wind energy analysis. This review includes the initial methods proposed in the 1940s which generally attempted only to estimate the long-term mean annual wind speed from a single reference station, and extends up to the present day and the most recent methods based on automatic learning techniques that consider several reference stations and which are not yet commonly used in the wind energy industry. This review covers more than 150 presentations at international wind energy congresses, papers published in IF journals in the field of renewable energies, commercial software containing MCP modules, PhD and Master's theses obtained from a variety of universities, books published by prestigious publishing houses as well as research studies undertaken in this field. The reading and examination of the above works has enabled the authors of the this paper to present not only a description of the linear, nonlinear and probabilistic transfer functions used by the different algorithms, but also the hypotheses on which they are based and the format in which they work (time series or frequency distributions). Also commented upon are the restrictions in the use of the MCP models, the uncertainty associated with them as well as the different reference data sources that have been used. All of which amounts to an interesting catalogue of MCP methods which we trust will be of use for renewable energy researchers, particularly in the field of wind energy.

Acknowledgements

The authors would like to express their gratitude to Dr. Peter Clive (Technical Development Officer, SgurrEnergy Ltd.), Dr. Petros Dellaportas (Department of Statistics, Athens University of Economics and Business), Mr. Matt Rebbeck (Chief Operating Officer, RES Australia Pty Ltd) and to Ms. Maura Di Ruscio (Administrative Assistant, EWEA) for their kindness and promptness in responding to requests for material which have allowed us to complete this paper. We would also like to thank Mistaya Engineering Inc. (WindoGrapher), EMD International A/S (WindPro), Sander+Partner GMBH (Mint), Enviroware (WindRose Pro), ReSoft Ltd (WindFarm) and GL Garrad Hassan (WindFarmer), for their permission to download the software they have developed. Our thanks also go to the State Meteorological Agency (Spanish initials: AEMET) of the Ministry of the Environment and Rural and Marine Environs of the Spanish Government and to the Canary Islands Technological Institute (Spanish initials: ITC) for providing us with the data series used for the examples presented in the figures and tables of this review.

References

[1] Carta JA. Wind power integration. In: Sayigh A, editor. Comprehensive renewable energy. Oxford: Elsevier; 2012. p. 569–622.
[2] Jain P. Wind energy engineering. 1st ed.. New York: McGraw-Hill; 2011.
[3] Bueno C, Carta JA. Wind powered pumped hydro storage systems, a means of increasing the penetration of renewable energy in the Canary Islands. Renewable and Sustainable Energy Reviews 2006;10:312–40.
[4] Velázquez S, Carta JA, Matías JM. Comparison between ANNs and linear MCP algorithms in the long-term estimation of the cost per kWh produced by a wind turbine at a candidate site. A case study in the Canary Islands. Applied Energy 2011;88:3869–81.
[5] EWEA. Wind energy – the facts. 1st ed. London: Earthscan; 2009.
[6] Hiester TR, Pennell WT. The siting handbook for large wind energy systems. 1st ed. New York: WindBook; 1981.
[7] Albers A, Klug H. High quality wind speed measurements for site assessment. DEWI Magazin 1999;15:6–16.
[8] Raftery P, Tindal A, Garrad A. Understanding the risks of financing wind farms. In: Watson R, editor. European wind energy conference. Dublin Castle: Irish Wind Energy Association; 1998. p. 77–81.
[9] Brower MC. Wind resource assessment. 1st ed. New Jersey: Wiley; 2012.
[10] Justus CG, Mani K, Mikhail AS. Interannual and month-to-moth variations of wind speed. Journal of Applied Meteorology 1979;18:913–20.
[11] Gerdes G, Strack M. Long-term correlation of wind measurement data. DEWI Magazin 1999;15:18–24.
[12] Ramsdell JV, Houston S, Wegley HL. Measurement strategies for estimating long-term average wind speeds. Solar Energy 1980;25:495–503.
[13] Landberg L, Myllerup L, Rathmann O, Petersen EL, Jørgensen BH, Niels BJ, et al. Wind resource estimation – an overview. Wind Energy 2003;6:261–71.
[14] Gasch R, Twele J. Wind power plants. 2nd ed. Berlin: Springer; 2012.
[15] Lackner MA, Rogers AL, Manwell JF. Uncertainty analysis in MCP-based wind resource assessment and energy production estimation. Journal of Solar Energy Engineering 2008;130:031006/1–10.
[16] Prasad RD, Bansal RC. Technologies and methods used in wind resource assessment. In: Zobaa AF, Bansal RC, editors. Handbook of renewable energy technology. Singapore: World Scientific Publishing Co. Pte. Ltd; 2011. p. 69–98.
[17] Derrick A. Development of the measure-correlate-predict strategy for site assessment. In: Clayton BR, editor. Fourteenth British wind energy conference. Nottingham: Mechanical Engineering Publications Ltd; 1992. p. 259–65.
[18] Palutikof JP, Kelly PM, Davies TD, Halliday JA. Impacts of spatial and temporal wind speed variability on wind energy output. Journal of Climate and Applied Meteorology 1987;26:1124–33.
[19] Langreder W. Wind resource and site assessment. In: Tong W, editor. Wind power generation and wind turbine design. MA: WIT Press; 2010. p. 49–87.
[20] Angelis-Dimakis A, Biberacher M, Dominguez J, Fiorese G, Gadocha S, Gnansounou E, et al. Methods and tools to evaluate the availability of renewable energy sources. Renewable and Sustainable Energy Reviews 2011;15:1182–200.

[21] Gass V, Strauss F, Schmidt J, Schmid E. Assessing the effect of wind power uncertainty on profitability. Renewable and Sustainable Energy Reviews 2011;15:2677–83.
[22] Addison JF, Hunter A, Bass J, Rebbeck M. A neural network version of the measure–correlate–predict algorithm for estimating wind energy yield. In: Proceedings of the 13th international congress and exhibition on condition monitoring and diagnostic engineering management, Houston, Texas; 3–8 December 2000. p. 917–922.
[23] Bowen AJ, Mortensen NG. Exploring the limits of WAsP: the wind atlas analysis and application program. In: Zervos A, Ehmann H, editors. European union wind energy conference. Bedford: Stephens HS and Associates; 1996. p. 584–7.
[24] Thøgersen ML, Nielsen P, Sørensen T, Svenningsen LU. An introduction to the MCP facilities in WindPRO. EMD International A/S; 2010 ⟨http:/help.emd.dk/knowledgebase/content/ReferenceManual/MCP.pdf⟩ [accessed on 22 March 2013].
[25] Windographer. Mistaya Engineering Inc. 109 Arbour Ridge Heights NW, Calgary AB T3G 3Z1, Canada ⟨http:/www.windographer.com/professional-edition⟩ [accessed on 22 March 2013].
[26] WindFarm ReSoft Ltd. Comwallis, Burycroft Road, Hook Norton, Banbury OX15 5PR, United Kingdom ⟨http:/resoft.co.uk/English/body_index.htm⟩ [accessed on 22 March 2013].
[27] WindFarmer ⟨http:/www.glgarradhassan.com/assets/img/content/MCP_module_brochure.pdf⟩ [accessed on 22 March 2013].
[28] Mint ⟨http:/www.sander-partner.com/en/products/mint-details.html⟩ [accessed on 22 March 2013].
[29] WindRose ⟨http:/www.windrose.gr/⟩ [accessed on 22 March 2013].
[30] WindLogics ⟨http:/www.windlogics.com/wp-content/uploads/2012/04/WindLogics2008-The-Long-Term-Wind-Resource-Comparing-Data-Sources-and-Techniques-for-Predicting-the-Performance-of-Wind-Plants-Part-2.pdf⟩ [accessed on 22 March 2013].
[31] Derrick A, Ravey I, Marti I, Glinou G, Pahlke T, Schwenk B, Brand A, Antoniou I, Frandsen S. A unified approach to the evaluation of site specific wind characteristics for use in both energy and load modeling of a potential wind turbine development site. In: Watson R, editor. Proceeding of the EWEC. Ireland: IWEA; 1997. p. 300–5.
[32] Sanz J. State-of-the-art of wind resource assessment ⟨http:/www.waudit-itn.eu/download.php?id=103&parent=79⟩ [accessed on 22 March 2013].
[33] Landberg L, Mortensen NG. A comparison of physical and statistical methods for estimating the wind resource at a site. In: Pitcher KF, editor. Proceedings of the 15th British wind energy association conference. Mechanical Engineering Publications Ltd.; 1993. p. 119–25.
[34] Bowen AJ, Mortensen NG. WAsP prediction errors due to site orography. Risø National Laboratory; 2004 ⟨http:/orbit.dtu.dk/fedora/objects/orbit:91202/datastreams/file_7711496/content⟩ [accessed on 22 March 2013].
[35] Khadem SK, Badger J, Ullah SM, Aditya SK, Ghosh HR, Hussain M.The effect of obstacles close to the anemometer mast located on a building on wind flow in the WAsP model, RETRUD 03, Nepal, 12–14 October 2003.
[36] Khadem SK, Hussain M. A pre-feasibility study of wind resources in Kutubdia Island, Bangladesh. Renewable Energy 2006;31:2329–41.
[37] Abbes M, Belhadj J. Wind resource estimation and wind park design in El-Kef region, Tunisia. Energy 2012;40:348–57.
[38] Xydis G. Wind-direction analysis in coastal mountainous sites: an experimental study within the Gulf of Corinth, Greece. Energy Conversion and Management 2012;64:157–69.
[39] Khan MJ, Iqbal MT. Wind energy resource map of Newfoundland. Renewable Energy 2004;29:1211–21.
[40] Manwell JF, Elkinton CN, Rogers AL, McGowan JG. Review of design conditions applicable to offshore wind energy systems in the United States. Renewable and Sustainable Energy Reviews 2007;11:210–34.
[41] Manwell JF, Rogers AL, McGowan JG, Bailey BH. An offshore wind resource assessment study for New England. Renewable Energy 2002;27:175–87.
[42] Lavagnini A, Sempreviva AM, Barthelmie RJ. Estimating wind energy potential offshore in Mediterranean areas. Wind Energy 2003;6:23–34.
[43] Oh KY, Kim JY, Lee JK, Ryu MS, Lee JS. An assessment of wind energy potential at the demonstration offshore wind farm in Korea. Energy 2012;46:555–63.
[44] Carta JA, Calero R, Colmenar A, Castro MA, Collado E. Renewable energy power plants. Electricity generation with renewable energies. 2nd ed. Madrid: Pearson; 2012 (in Spanish).
[45] Thuman C, Schnitzer M, Johnson P. Quantifying the accuracy of the use of measure-correlate-predict methodology for long-term solar resource estimates. In: Fellows C, editor. World Renewable Energy Forum (WREF), 4. Colorado: American Solar Energy Society; 2012. p. 2639–43.
[46] Fontaine A, Armstrong P. Uncertainty analysis in energy yield assessment. In: Proceedings of the European wind energy conference, Milan, Italy; 7–10 May 2007.
[47] Rogers AL, Rogers JE, Manwell JF. Uncertainties in results of measure-correlate-predict analyses. In: Proceedings of the European wind energy conference & exhibition,, Athens, Greece; 27 February–2 March 2006.
[48] Faulkner S. Quantifying the uncertainties of correlation-prediction (MCP). In: Proceedings of the New Zealand wind energy conference and exhibition, Palmerston North, New Zealand; 29–31 March 2010.
[49] Taylor M, Mackiewicz P, Brower MC, Markus M. An analysis of wind resource uncertainty in energy production estimates. In: Proceedings of the European wind energy conference & exhibition, London, UK, 22–25 November 2004.
[50] Anderson M. MCP errors ⟨http:/www.res-group.com/media/234588/mcp%20errors.pdf⟩ [accessed on 22 March 2013].
[51] Derrick A. Uncertainty: the classical approach. In: AWEA wind resource & project assessment workshop, Minneapolis, MN, USA; 30 September–1 October 2009.
[52] Ramli SC. Uncertainty in the application of the measure-correlate-predict method in wind resource assessment. In: Proceedings of the EWEA offshore 2011, Amsterdam, The Netherlands; 29 November–1 December 2011.
[53] Hume-Wright L, Lee M, Skea A. Accounting for uncertainty unquantified in MCP. In: Proceedings of the EWEA annual conference and exhibition, Messe Wien, Vienna, Austria; 4–7 February 2013.
[54] Lloyd W. Wind resource assessment using measure-correlate-predict techniques, MSc thesis, CREST, Loughborough University; 1995.
[55] Rebbeck MA Comparison of measure-correlate-predict techniques for wind resource assessment. MSc thesis, Department of Electronic and Electrical Engineering, Loughborough University, Loughborough, Leicestershire; September 1996.
[56] Sheppard CJR. Analysis of the measure-correlate-predict methodology for wind resource assessment. Thesis presented to The Faculty of Humboldt State University. November; 2009.
[57] Saengyuenyongpipat P. Demonstrating measure-correlate-predict algorithms for estimation of wind resources in Central Finland. Master's thesis, University of Jyväskylä, Department of Physics, Master's Degree Programme in Renewable Energy; March 2010.
[58] ⟨http:/help.emd.dk/knowledgebase/content/WindPRO2.8/11-UK_WindPRO2.8_MCP.pdf⟩ [accessed on 22 March 2013].
[59] AWS Truewind ⟨http:/www.awstruepower.com/wp-content/media/2010/05/ResearchNoteVol1.pdf⟩ [accessed on 22 March 2013].
[60] Mina G, Clive P. Assessing the influence of neighbouring wind farms on one another. In: Proceedings of the EWEA annual conference and exhibition, Bella Center, Copenhagen, Denmark; 16–19 April 2012.
[61] Probst O, Cárdenas D. State of the art and trends in wind resource assessment. Energies 2010;3:1087–141.
[62] Oliver A, Zarling K. Time of Day Correlations for Improved Wind Speed Predictions. In: Proceedings of the AWEA 2009 windpower conference and exhibition, Chicago IL, USA; 30 April–8 May 2009.
[63] Emeis S. Wind energy meteorology. 1st ed. New York: Springer; 2013.
[64] Palma JMLM, Castro FA, Ribeiro LF, Rodrigues AH, Pinto AP. Linear and nonlinear models in wind resource assessment and wind turbine micro-siting in complex terrain. Journal of Wind Engineering and Industrial Aerodynamics 2008;96:2308–26.
[65] Anderson M.A Review of MCP techniques ⟨http:/www.res-group.com/media/234585/a%20review%20of%20mcp%20techniques.pdf⟩ [accessed on 22 March 2013].
[66] Draper NR, Smith H. Applied regression analysis. 3rd ed. New York: John Wiley & Sons, Inc; 1998.
[67] Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting. Methods and applications. 3rd ed. New York: Wiley; 1998.
[68] Brandimarte P. Quantitative methods: an introduction for business management. 1st ed. New Jersey: Wiley; 2011.
[69] Wilks DS. Statistical methods in the atmospheric sciences. 3rd ed. USA: Academic Press; 2011.
[70] McKenzie J, Clive P, Bulté H, Chindurza I. Correlation & correlation measurement in measure-correlate-predict techniques. In: Global Wind Power, Beijing, China; 29–31 October 2008.
[71] Canavos G. Applied probability statistical methods. 1st ed. New York: Little Brown & Company; 1998.
[72] Ayotte KW, Davy RJ, Coppin PA. A simple temporal and spatial analysis of flow in complex terrain in the context of wind energy modeling. Boundary-Layer Meteorology 2001;98:275–95.
[73] Früh WG. Long-term wind resource and uncertainty estimation using wind records from Scotland as example. Renewable Energy 2013;50:1014–26.
[74] Nielsen M, Landberg L, Mortensen NG, Barthelmie RJ, Joensen A. Application of the measure-correlate-predict approach for wind resources assessment. In: Helm P, Zervos A, editors. European wind energy conference. Wind energy for the new millennium. Munich: WIP-Renewable Energies; 2001. p. 773–6.
[75] Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 2nd ed.Numerical recipes in Fortran 77, vol. 1. New York: Cambridge University Press; 1992.
[76] Bass J. MCP: Pitfalls & common mistakes. In: AWEA wind resource & project assessment workshop, Minneapolis, MN, USA; 30 September–1 October 2009.
[77] Oliver A, Zarling K. The effect of seasonality on wind speed prediction bias in the plains. In: Proceedings of the AWEA 2010 windpower conference and exhibition, Dallas Texas, USA; May 23–25, 2010.
[78] Freedman JM, Waight KT, Duffy PB. Does climate change threaten wind resources? North American Windpower 2009 ([May 16]).
[79] Klink K. Trends in monthly maximum and minimum surface wind speeds in the coterminous United States, 1961 to 1990. Climate Research 1999;13:193–205.
[80] Smits A, Klein-Tank AMG, Können GP. Trends in storminess over the Netherlands, 1962–2002. International Journal of Climatology 2005;25:1331–44.
[81] Vautard R, Cattiaux J, Yiou P, Thépaut JN, Ciais P. Northern Hemisphere atmospheric stilling partly attributed to an increase in surface roughness. Nature Geoscience 2010;3:756–61.

[82] AWS TruePower 〈http:/www.awstruepower.com/2010/11/aws-truepower-questions-reported-diminishing-wind-speeds/〉 [accessed on 22 March 2013].
[83] Breslow PB, Sailor DJ. Vulnerability of wind power resources to climate change in the continental United States. Renewable Energy 2002;27:585–98.
[84] Hazlett M. Climate change could have major impacts on wind resources. North American Windpower 2011 ([January 04].).
[85] Sailor DJ, Smith M, Hart M. Climate change implications for wind power resources in the Northwest United States. Renewable Energy 2008;33:2393–406.
[86] Watson S, Kritharas P. Long term wind speed variability in the UK. In: Proceedings of the EWEA annual conference and exhibition, Bella Center, Copenhagen, Denmark; 16–19 April 2012.
[87] Pryor SC, Barthelmie RJ. Climate change impacts on wind energy: a review. Renewable and Sustainable Energy Reviews 2010;14:430–7.
[88] Durre I, Vose RS, Wuertz DB. Overview of the integrated global radiosonde archive. Journal of Climate 2006;19:53–68.
[89] Taylor M, Freedman J, Waight K, Brower M. Using simulated wind data from a mesoscale model in MCP. In: AWEA windpower, Chicago, IL, USA; 30 April–8 May 2009.
[90] Brower MC. The use of NCEP/NCAR reanalysis data in MCP. In: Proceedings of the European wind energy conference & exhibition, Athens, Greece; 27 February–2 March 2006.
[91] Liléo S, Petrik O. Investigation on the use of NCEP/NCAR, MERRA and NCEP/CFSR reanalysis data in wind resource analysis. In: Proceedings of the European wind energy conference & exhibition, Brussels, Belgium, 14–17 March 2011.
[92] Pinto C, Guedes R, Pinto P, Ferreira M. NCEP/NCAR reanalysis for the Portuguese mainland. In: Proceedings of the European wind energy conference & exhibition, Athens, Greece; 27 February–2 March 2006.
[93] NCEP/NCAR. 〈http:/www.esrl.noaa.gov/psd/data/reanalysis/reanalysis.shtml〉 [accessed on 22 March 2013].
[94] ECMWF 〈http:/www.ecmwf.int/〉 [accessed on 22 March 2013].
[95] MERRA http:/merra-reanalysis.blogspot.com.es/ [accessed on 22 March 2013].
[96] NCEP/CFSR 〈https:/climatedataguide.ucar.edu/reanalysis/climate-forecast-system-reanalysis-cfsr〉 [accessed on 22 March 2013].
[97] Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, Dool H, Jenne R. The NCEP/NCAR 50-Year reanalysis: monthly means CD-ROM and documentation. Bulletin of the American Meteorological Society 2001;82:247–68.
[98] King C, Hurley B. The SpeedSort, DynaSort and Scatter wind correlation methods. Wind Engineering 2005;29:217–41.
[99] Beltrán J, Cosculluela L, Pueyo C, Melero JJ. Comparison of measure-correlate-predict methods in wind resource assessments. In: Proceedings of the European wind energy conference, Warsaw, Poland; 20–23 April 2010.
[100] King C, Hurley B. The moulded site data (MSD) wind correlation method; description and assessment. Wind Engineering 2004;28:649–66.
[101] Woods JC, Watson SJ. A new matrix method of predicting long-term wind roses with MCP. Journal of Wind Engineering and Industrial Aerodynamics 1997;66:85–94.
[102] Velázquez S, Carta JA, Matías JM. Influence of the input layer signals of ANNs on wind power estimation for a target site: a case study. Renewable and Sustainable Energy Reviews 2011;15:1556–66.
[103] Carta JA, Velázquez S, Matías JM. Use of Bayesian networks classifiers for long- term mean wind turbine energy output estimation at a potential wind energy conversion site. Energy Conversion & Management 2011;52:1137–49.
[104] Bass JH, Rebbeck M, Landberg L, Cabré M, Hunter A. An improved measure-correlate-predict algorithm for the prediction of the long term wind climate in regions of complex environment. Joule Project JOR3-CT98-0295. Final Report. 〈http:/www.res-group.com/media/234621/jor3-ct98-0295-finalreport.pdf〉 [accessed on 22 March 2013].
[105] Thøgersen ML, Motta M, Sørensen T, Nielsen P. Measure-correlate-predict methods: case studies and software implementation. In: Proceedings of the European wind energy conference, Milan, Italy; 7–10 May 2007.
[106] Rogers AL, Rogers JW, Manwell JF. Comparison of the performance of four measure–correlate–predict algorithms. Journal of Wind Engineering and Industrial Aerodynamics 2005;93:243–64.
[107] Riedel V, Strack M, Waldl H. Robust approximation of functional relationships between meteorological data: alternative measure-correlate-predict algorithms. In: Helm P, Zervos A, editors. European wind energy conference. Wind energy for the new millennium. Munich: WIP-Renewable Energies; 2001. p. 806–9.
[108] 〈http:/cordis.europa.eu/search/index.cfm?fuseaction=proj.document&PJ_RCN=3055235〉.
[109] Putnam PC. Power from the wind. 1st ed. New York: Van Nostrand Reinhold Company; 1948.
[110] Koeppl GW. Putnam's power from the wind. 2nd ed. New York: Van Nostrand Reinhold Company; 1982.
[111] Corotis RB. Stochastic modeling of site wind characteristics. Northwestern University, Department of Civil Engineering. ERDA Contract No. EY-76-S-06-2342. Final Report; 1977.
[112] Daniels PA, Schroede TA. Siting large wind turbines in Hawaii. Wind Engineering 1988;12:302–30.
[113] Barchet WR, Davis WE. Estimating long-term mean winds from short-term wind data. Pacific Northwest Labs., Richland, WA. Report Number PNL-4785; August 1987 〈www.ntis.gov〉.
[114] Carta JA, González J. Self-sufficient energy supply for isolated communities: wind-diesel systems in the Canary Islands. The Energy Journal 2001;22:115–45.
[115] Conrad V, Pollack LW. Methods in climatology. Cambridge (MA). 2nd ed. Harvard University Press; 1962.
[116] Barros VR, Estevan EA. On the evaluation of wind power from short wind records. Journal of Climate and Applied Meteorology 1983;22:1116–23.
[117] Tallhaug L, Nygaard TA. The potential of wind energy in Sor-Trondelag, Norway. In: Garrad AD, Palz W, Screller S, editors. European Community Wind Energy Conference. Bedford: Stephens HS and Associates; 1993. p. 87–90.
[118] Harstveit K. Estimating long-term wind distributions from short-term data set using a reference station. In: Proceedings of the European wind energy conference & exhibition, London, UK; 22–25 November, 2004.
[119] Nygaard TA. Estimating expected energy capture at potential wind turbine sites in Norway. Journal of Wind Engineering and Industrial Aerodynamics 1992;39:385–93.
[120] Barchet WR. A weather pattern climatology of the Great Plains and the related wind regime. Pacific Northwest Labs., Richland, WA. Report Number PNL-4330; November 1982. 〈www.ntis.gov〉.
[121] Barchet WR, Davis WE. A weather pattern climatology of the United States. Pacific Northwest Labs., Richland, WA. Report Number PNL-4889; January 1984 〈www.ntis.gov〉.
[122] Derrick A. Development of the measure-correlate-predict strategy for site assessment. In: Garrad AD, Palz W, Screller S, editors. European Community wind energy conference. Bedford: Stephens HS and Associates; 1993. p. 681–5.
[123] Bardsley WE, Manly BFJ. Regression-based estimation of long-term mean and variance of wind speed at potential aerogenerator sites. Journal of Applied Meteorology 1983;22:323–7.
[124] Prasad RD, Bansal RC, Sauturaga M. Some of the design and methodology considerations in wind resource assessment. IET Renewable Power Generation 2009;3:53–64.
[125] Romo A, Amezcua J, Probst O. Validation of three new measure-correlate-predict models for the long-term prospection of the wind resource. Journal of Renewable Sustainable Energy 2011;3:023105, http://dx.doi.org/10.1063/1.3574447.
[126] Zhang J, Chowdhury S, Messac A, Castillo L. A hybrid measure-correlate-predict method for wind resource assessment. In: Proceedings of the ASME 2012 6th international conference on energy sustainability & 10th fuel cell science, engineering and technology conference ESFuelCell2012, San Diego, CA, USA; 23–26 July 2012.
[127] Kwak G, Yang KB, Ko KN, Huh JCA. Comparison of measure-correlate-predict method in Jeju Island. Journal of Wind Energy 2010;1:79–86.
[128] LeBlanc M, Schoborg D, Cox S, Haché A, Tindal A. Is a Non-linear MCP method a useful tool for North American wind regimes? In: Proceedings of the AWEA 2009 windpower conference and exhibition, Chicago, IL USA; 30 April–8 May, 2009.
[129] Hanslian D. Two measure-correlate-predict methods and their performance. In: Proceedings of the European wind energy conference & exhibition, Brussels, Belgium; 31 March–3 April; 2008.
[130] McKenzie J, Clive P, Chindurza I, Bulté H. Considering the correlation in measure-correlate-predict techniques. In: Sayigh A, editor. Proceedings of the World Renewable energy congress; 2008, p. 2287–229.
[131] McKenzie J, Clive P, Bulté H, Chindurza I. Considering the correlate in measure-correlate-predict techniques. In: International conference on renewable energy, Busan, South Korea; 13–17 October 2008.
[132] Joensen A, Landberg L, Madsen HA. New measure-correlate-predict approach for resource assessment. In: Petersen EL, Jensen PH, Rave K, Helm P, Ehmann H, editors. Proceeding of the EWEC. London: James and James Science Publisher; 1999. p. 1157–60.
[133] Clive PJM. Non-linearity in MCP with Weibull distributed wind speeds. Wind Engineering 2008;32:319–24.
[134] Sreevalsan E, Das SS, Sasikumar R, Ramesh MP. Wind farm site assessment using measure–correlate–predict (MCP) analysis. Wind Engineering 2007;31:111–6.
[135] van Lieshout P. Improvements in AEP calculations using IEC 61400.Wind Tech International 2010; May:1–6.
[136] Mortimer AA. A new correlation/prediction method for potential wind farm sites. In: Elliot G, editor. Sixteenth British wind energy conference. London: Mechanical Engineering Publications Ltd; 1994. p. 349–52.
[137] Salmon JR, Walmsley JL. A two-site correlation model for wind speed, direction and energy estimates. Journal of Wind Engineering and Industrial Aerodynamics 1999;79:233–68.
[138] Walmsley JL, Bagg DL. A method of correlating wind data between two stations with application to the Alberta oil sands. Atmosphere-Ocean 1978;16:333–47.
[139] García-Rojo R. Algorithm for the estimation of the long-term wind climate at a meteorological mast using a joint probabilistic approach. Wind Engineering 2004;28:213–24.
[140] Casella L. Long term correction when varying correlation: a general rule using a joint probabilistic approach. In: Proceedings of the European wind energy conference & exhibition, Brussels, Belgium; 14–17 March 2011.
[141] Carta JA, Velázquez S. A new probabilistic method to estimate the long-term wind speed characteristics at a potential wind energy conversion site. Energy 2011;36:2671–85.

[142] Lambert T, Grue A. The matrix time series method for MCP. In: Proceedings of the AWEA 2012 windpower conference and exhibition, Atlanta GA, USA; 3–6 June 2012.
[143] Romo A, Amezcua J, Probst O. Validation of a new MCP method using data from texan coastal sites. In: Proceedings of the AWEA 2009 windpower conference and exhibition, Chicago IL, USA; 30 April–8 May, 2009.
[144] Carta JA, Ramírez P, Velázquez S. A review of wind speed probability distributions used in wind energy analysis: case studies in the Canary Islands. Renewable and Sustainable Energy Reviews 2009;13:933–55.
[145] Carta JA, Bueno C, Ramírez P. Statistical modelling of directional wind speeds using mixtures of von Mises distributions: a case study. Energy Conversion and Management 2008;49:897–907.
[146] Carta JA, Ramírez P, Bueno C. A joint probability density function of wind speed and direction for wind energy analysis. Energy Conversion and Management 2008;48:1309–20.
[147] Ross SM. Simulation. 5th ed. CA: Academic Press; 2013.
[148] Lolli B, Gasperini P. A comparison among general orthogonal regression methods applied to earthquake magnitude conversions. Geophysical Journal International 2012;190:1135–51.
[149] Fuller WA, editor. Measurement error models. 1st ed.. New York: John Wiley & Sons, Inc; 1987.
[150] York D. Least-squares fitting of a straight line. Canadian Journal of Physics 1966;44:1079–86.
[151] Carr JR. Orthogonal regression a teaching perspective. International Journal of Mathematical Education in Science and Technology 2012;43:134–43.
[152] Castellaro S, Bormann P. Performance of different regression procedures on the magnitude conversion problem. Bulletin of the Seismological Society of America 2007;97:1167–75.
[153] Carroll RJ, Ruppert D. The use and misuse of orthogonal regression in linear errors-in-variables models. The American Statistician 1996;50:1–6.
[154] Krystek M, Anton M. A weighted total least-squares algorithm for fitting a straight line. Measurement Science and Technology 2007;22:3438–42.
[155] Krystek M, Anton M. A least-squares algorithm for fitting data points with mutually correlated coordinates to a straight line. Measurement Science and Technology 2011;18(035–101):9p (p).
[156] Koenker R, Bassett G. Regression quantiles. Econometrica 1978;46:33–50.
[157] Hao L, Naiman DQ. Quantile regression. 1st ed. California: SAGE Publications Inc; 2007.
[158] Beltrán J, Llombart A, Guerrero JJ. A bin method with data range selection for detection of nacelle anemometers faults. In: Proceedings of the European wind energy conference & exhibition, Marseille, France; 16–19 March 2009.
[159] IEC 2005. Wind turbines – Part 12-1: Power performance measurements of electricity producing wind turbines. IEC 61400-12-1 International Standard.
[160] Achberger C. Estimation of local near-surface wind conditions – a comparison of WASP and regression based techniques. Meteorological Applications 2002;9:211–21.
[161] Hanson B, Klink K, Matsuura K, Robeson SM, Willmott CJ. Vector correlation: review, exposition, and geographic application. Annals of the Association of American Geographers 1992;82:103–66.
[162] Vermuelen PEJ, Marijanyan A, Abrahamyan A, den Boon JH. Application of matrix MCP analysis in mountainous Armenia. In: Helm P, Zervos A, editors. European wind energy conference. Wind energy for the new millennium. Munich: WIP-Renewable Energies; 2001. p. 737–40.
[163] Deane JP, Moehrlen CS, McKeogh EJ. Wind data analysis. In: Helm P, Zervos A, editors. European wind energy conference. Wind energy for the new millennium. Munich: WIP-Renewable Energies; 2001. p. 846–9.
[164] Ramírez P, Carta JA. Influence of the data sampling interval in the estimation of the parameters of the Weibull wind speed probability density distribution: a case study. Energy Conversion and Management 2005;46:2419–38.
[165] Carta JA, Ramírez P. Use of finite mixture distribution models in the analysis of wind energy in the Canarian Archipelago. Energy Conversion and Management 2007;48:281–91.
[166] Carta JA, Ramírez P. Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions. Renewable Energy 2007;32:518–31.
[167] Jaramillo OA, Borja MA. Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case. Renewable Energy 2004;29:1613–30.
[168] Hunt K, Nason GP. Wind speed modelling and short-term prediction using wavelets. Wind Engineering 2001;25:55–61.
[169] Kotz S, van Dorp JR. Beyond beta. NJ: World Scientific Publishing Co Pte Ltd;; 2004.
[170] Efron B, Tibshirani RJ. An introduction to the bootstrap. 1st ed. New York: Chapman & Hall/CRC; .
[171] Modica G, Poggiolini LA. First course in probability and Markov chains. 1st ed. UK: Wiley; 2013.
[172] Sahin AD, Sen Z. First-order Markov chain approach to wind speed modelling. Journal of Wind Engineering and Industrial Aerodynamics 2001;89:263–9.
[173] Shamshad A, Bawadi MA, Wan Hussin WMA, Majid TA, Sanusi SAM. First and second order Markov chain models for synthetic generation of wind speed time series. Energy 2005;30:693–708.
[174] Alpaydin E. Introduction to machine learning. 2nd ed. Massachusetts: The MIT Press; 2010.
[175] Johnson RA, Evans JW, Green DW. Some bivariate distributions for modeling the strength properties of lumber, Department of Agriculture, Forest Service, Forest Products Laboratory; 1999. 11 p 〈http:/www.fpl.fs.fed.us/documnts/fplrp/fplrp575.pdf〉 [accessed on 22 March 2013].
[176] Walls L, Kline J, Kline Z. Long-term wind speed estimates from short-term data: so many ways to get it wrong! In: Proceedings of the AWEA wind resource assessment workshop 2010, 14 September, Oklahoma City, Oklahoma, USA; 2010.
[177] Oh KY, Kim JY, Lee JS. A study on the reduction of uncertainty in estimations of long term wind resources, using the complementary MCP (Measure-Correlate-Predict) Technique. In: Proceedings of the European wind energy conference & exhibition, Brussels, Belgium; 14–17 March 2011.
[178] Casella L. Improving long-term wind speed assessment using joint probability functions applied to three wind data sets. Wind Engineering 2012;36:473–84.
[179] Barros VR, Schmidt IG. On extension of climatic series from short records. Journal Applied Meteorology 1988;27:325–35.
[180] Zaphiropoulos Y, Dellaportas P, Morfiadakis E, Clinou G. Prediction of wind speed and direction at a potential site. Wind Engineering 1999;23:167–75.
[181] Haslett J, Raftery A. Space-time modeling with long-memory dependence: assessing Ireland's wind power resource. Applied Statistics 1989;38:1–50.
[182] Glinou G, Morfiadakis E, Zaphiropoulos Y, Dellaportas PA. Statistical approach to wind potential assessment using multivariate ARFIMA modelling. In: Petersen EL, Jensen PH, Rave K, Helm P, Ehmann H, editors. Proceeding of the EWEC. London: James and James Science Publisher; 1999. p. 1138–41.
[183] Carlin J, Haslett J. The probability distribution of wind power from a dispersed array of wind turbine generators. Journal of Applied Meteorology 1982;21:303–13.
[184] Denison DGT, Dellaportas P, Mallick BK. Wind speed prediction in a complex terrain. Environmetrics 2001;12:499–515.
[185] Colak I, Sagiroglu S, Yesilbudak M. Data mining and wind power prediction: a literature review. Renewable Energy 2012;46:241–7.
[186] Izenman AJ. Modern multivariate statistical techniques. 1st ed. New York: Springer; 2008.
[187] Kalogirou SA. Artificial neural networks in renewable energy systems applications: a review. Renewable and Sustainable Energy Reviews 2001;5:373–401.
[188] Addison JF, Hunter A, Bass J, Rebbeck M. A neural network version of the measure correlate predict algorithm for estimating wind energy yield. In: Proceedings of 13th international congress on condition monitoring and diagnostic engineering, Houston, TX; 2000. p. 917–22.
[189] Bechrakis DA, Deane JP, McKeogh EJ. Wind resource assessment of an area using short term data correlated to a long term data set. Solar Energy 2004;76:725–32.
[190] Öztopal A. Artificial neural network approach to spatial estimation of wind velocity. Energy Conversion and Management 2006;47:395–406.
[191] Bilgili M, Sahin B, Yasar A. Application of artificial neural networks for the wind speed prediction of target station using reference stations data. Renewable Energy 2007;32:2350–60.
[192] Monfared M, Rastegar H, Kojabadi HM. A new strategy for wind speed forecasting using artificial intelligent methods. Renewable Energy 2009;34:845–8.
[193] Lopez P, Velo R, Maseda F. Effect of direction on wind speed estimation in complex terrain using neural networks. Renewable Energy 2008;33:2266–72.
[194] Patanè D, Benso M, Hernández C,. de La Blanca F, López C. Long term wind resource assessment by means of multivariate cross-correlation analysis. In: Proceedings of the European wind energy conference & exhibition, Brussels, Belgium; 14–17 March 2011.
[195] Zhang S, Jin J. Computation of special functions. 1st ed. New York: Wiley; 1996.